Traducción de
FRANCISCO DIAZ KLAASSEN

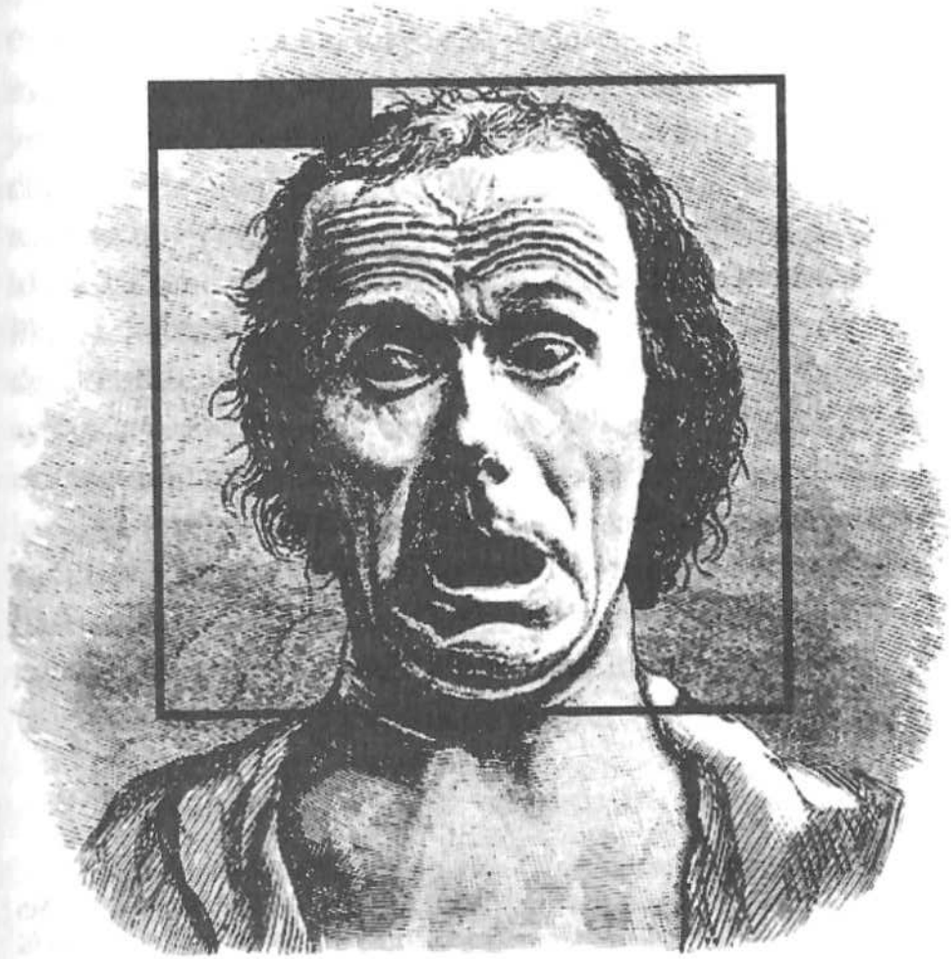
KATE CRAWFORD

ATLAS DE INTELIGENCIA ARTIFICIAL

Poder, política y costos planetarios

FONDO DE CULTURA ECONÓMICA
MÉXICO - ARGENTINA - BRASIL - CHILE - COLOMBIA - ECUADOR - ESPAÑA
ESTADOS UNIDOS DE AMÉRICA - GUATEMALA - PERÚ - VENEZUELA

V. Las emociones



En un remoto puesto de avanzada de las zonas montañosas de Papúa Nueva Guinea, un joven psicólogo estadounidense llamado Paul Ekman llegó acompañado de una colección de tarjetas y una nueva teoría.[1] Era el año 1967 y Ekman había escuchado que el pueblo fore de Okapa estaba tan aislado del ancho mundo que pensó que eso los volvería sus sujetos de prueba ideales. Como muchos investigadores occidentales antes que él, Ekman había llegado a Papúa Nueva Guinea para recabar datos de la comunidad indígena. Reunía pruebas para reforzar una controvertida hipótesis: que todos los seres humanos exhiben un pequeño número de emociones o afectos que son naturales, innatos, transculturales e idénticos en todo el mundo. Aunque esa afirmación sigue siendo poco convincente, sus consecuencias han tenido un gran alcance: los presupuestos de Ekman sobre las emociones se han convertido en una industria creciente que vale más de 17.000 millones de dólares.[2] Esta es la historia de

[1] Agradecimientos especiales a Alex Campolo, mi asistente de investigación e interlocutor en este capítulo, también por su investigación sobre Ekman y la historia de las emociones.

[2] "Emotion Detection and Recognition (EDR) Market Size to Surpass 18%+ CAGR 2020 to 2027", en *MarketWatch,* 5 de octubre de 2020, y Oscar Schwartz,

cómo el reconocimiento emocional se volvió parte de la inteligencia artificial (IA) y los problemas que esto conlleva.

En los trópicos de Okapa, guiado por el investigador médico Daniel Carleton Gajdusek y el antropólogo E. Richard Sorenson, Ekman esperaba conducir experimentos para evaluar la manera en que los fore reconocían emociones transmitidas por expresiones faciales. Puesto que los fore tenían un contacto mínimo con Occidente y los medios de comunicación, Ekman teorizaba que su reconocimiento y exhibición de expresiones básicas probaría que tales expresiones eran universales. Sus métodos eran simples: les mostraría tarjetas de expresiones faciales y observaría si, al verlas, describían las emociones. En sus propias palabras: "Todo lo que estaba haciendo era mostrarles imágenes divertidas".[3]

Pero Ekman no tenía ningún conocimiento de la historia, el lenguaje, la cultura y la política de los fore. Sus intentos por realizar sus experimentos con tarjetas usando traductores fracasaron; él y sus sujetos quedaban agotados por el proceso, que describió como sacar dientes sin anestesia.[4] Ekman abandonó Papúa Nueva Guinea frustrado por este primer intento de investigación transcultural sobre las expresiones emocionales. Pero esto solo sería el principio.

Las herramientas de reconocimiento emocional actuales se pueden encontrar en los sistemas de seguridad nacionales y en los aeropuertos, en la educación y en *startups* para la contratación de empleados, desde sistemas que pretenden detectar enfermedades psiquiátricas hasta programas de vigilancia que aseguran poder predecir la violencia. Si observamos la historia de cómo surgió la detección de emociones automática, podemos entender que sus métodos hayan planteado tanto preocupaciones éticas como dudas científicas. Como veremos, la afirmación de que el estado interior emocional de alguien se puede evaluar con precisión analizando su rostro se basa en una evidencia inestable.[5] De hecho, una revisión exhaustiva de la bibliografía científica disponible sobre la inferencia de emociones a partir de movimientos faciales, publicada en 2019, fue concluyente: *no hay evidencia confiable* de que puedes predecir con precisión el estado emocional de una persona a partir de su rostro.[6]

¿Cómo fue, entonces, que esta colección de afirmaciones y metodologías experimentales controvertidas se convirtió en el enfoque que impulsa muchas partes de la industria de la IA emocional? ¿Por qué la idea de que hay un pequeño conjunto de emociones universales que pueden interpretarse fácilmente a partir del rostro se volvió tan aceptada en el campo de la IA, a pesar de la existencia de considerable evidencia que prueba lo contrario? Para entender eso, es necesario rastrear el desarrollo de esas ideas, mucho antes de que las herramientas de reconocimiento emocional de la IA fueran integradas a la infraestructura de la vida cotidiana.

"Don't Look Now: Why You Should Be Worried about Machines Reading Your Emotions", en *The Guardian,* 6 de marzo de 2019.

[3] Miyoko Ohtake, "Psychologist Paul Ekman Delights at Exploratorium", en *Wired,* 28 de enero de 2008, disponible en línea: <https://www.wired.com/2008/01/psychologist-pa/>.

[4] Paul Ekman, *Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life,* Nueva York, Times, 2003, p. 7.

[5] Para una descripción general de los investigadores que han encontrado fallas en la afirmación de la universalidad de las expresiones emocionales y el que la IA pueda predecirlas, véase Douglas Heaven, "Why Faces Don't Always Tell the Truth about Feelings", en *Nature,* 26 de febrero de 2020, disponible en línea: <https://www.nature.com/articles/d41586-020-00507-5>.

[6] Lisa Feldman Barrett *et al.,* "Emotional Expressions Reconsidered: Challenges to Inferring Emotion from Human Facial Movements", en *Psychological Science in the Public Interest,* vol. 20, núm. 1, 2019, pp. 1-68, disponible en línea: <https://doi.org/10.1177/1529100619832930>

Ekman es solo una de las muchas personas que han contribuido a las teorías detrás del reconocimiento emocional. Pero la rica y sorprendente historia de su investigación ilumina algunas de las complejas fuerzas que impulsan el campo. Su trabajo está conectado, en primer lugar, con la financiación, por parte de las agencias de inteligencia de Estados Unidos, de las ciencias humanas durante la Guerra Fría, pero también con el trabajo fundamental en el campo de la visión artificial de los programas de seguridad posteriores al 11 de septiembre empleados para identificar a los terroristas, y así hasta llegar a la moda actual del reconocimiento emocional basado en la IA. Esta es una crónica que combina ideologías, políticas económicas y otras basadas en el miedo, así como el deseo de extraer más información sobre las personas de la que ellas están dispuestas a dar.

Profetas emocionales: cuando los sentimientos cuestan dinero

Para los ejércitos, las corporaciones, las agencias de inteligencia y las fuerzas policiales del mundo, la idea del reconocimiento emocional automático es tan llamativa como lucrativa. Promete filtrar, de manera confiable, al amigo del enemigo, distinguir mentiras de verdades y usar los instrumentos de la ciencia para adentrarse en mundos interiores.

Las compañías tecnológicas han capturado volúmenes inmensos de imágenes superficiales de expresiones humanas, incluidos miles de millones de selfis de Instagram, retratos de Pinterest, videos de TikTok y fotografías de Flickr. Una de las muchas cosas que la profusión de imágenes hizo posible fue el intento de extraer, usando el aprendizaje automático, la llamada verdad escondida de los estados emocionales interiores. El reconocimiento emocional se está integrando en muchas plataformas de reconocimiento facial, desde las compañías tecnológicas más grandes hasta las pequeñas *startups*. Mientras que el reconocimiento facial busca identificar a un individuo *particular,* la detección emocional apunta a detectar y clasificar emociones analizando *cualquier* rostro. Puede que estos sistemas no estén haciendo lo que pretenden hacer, pero pueden ser agentes poderosos a la hora de influir comportamientos y entrenar personas para que se desempeñen de maneras reconocibles. Estos sistemas ya están jugando un papel en delinear el comportamiento de las personas y la manera en que las instituciones sociales operan, a pesar de la falta de evidencia científica sustancial que diga que funcionan.

Hoy en día se utilizan ampliamente los sistemas automatizados de detección emocional, en especial en la contratación. Una *startup* de Londres llamada Human usa el reconocimiento emocional para analizar entrevistas filmadas de candidatos a puestos de trabajo. De acuerdo con un informe del *Financial Times,* "la compañía asegura que puede detectar las expresiones emocionales de los posibles candidatos y relacionarlas con rasgos de personalidad"; la compañía luego puntúa a los sujetos en rasgos de personalidad como honestidad o pasión por el trabajo.[7] HireVue, la compañía de contratación por medio de IA, que tiene entre sus clientes a Goldman Sachs, Intel y Unilever, usa aprendizaje automático para evaluar pistas faciales que les permitan inferir la idoneidad de una persona para un trabajo. En el año 2014, la compañía lanzó su sistema de IA para extraer microexpresiones, tonos de voz y otras variables a partir de entrevistas de trabajo filmadas, que usaron para comparar a los

[7] Patricia Nilsson, "How AI Helps Recruiters Track Jobseekers' Emotions", en *Financial Times,* 28 de febrero de 2018, disponible en línea: <https://www.ft.com/content/e2e85644-05be-11e8-9650-9c0ad2d7c5b5>.

candidatos con los trabajadores que tenían menor desempeño de la compañía.[8]

En enero de 2016, Apple compró la *startup* Emotient, que aseguraba haber producido un *software* capaz de detectar emociones a partir de imágenes de rostros.[9] Emotient surgió de una investigación académica llevada a cabo en la University of California San Diego, y es una de las muchas *startups* que trabajan en esta área.[10] Quizá la más grande de todas es Affectiva, una compañía con sede en Boston que emergió del trabajo académico realizado en el Massachusetts Institute of Technology (MIT). Aquí, Rosalind Picard y sus colegas formaban parte de un campo emergente más amplio, llamado computación afectiva, que describe la computación que "se relaciona con, surge de o influye deliberadamente en las emociones y otros fenómenos afectivos".[11]

Affectiva codifica una variedad de aplicaciones relacionadas con las emociones, usando principalmente técnicas de aprendizaje profundo, que van desde detectar a conductores distraídos o "arriesgados" en las carreteras hasta medir las respuestas emocionales de los consumidores frente a la publicidad. La compañía ha construido lo que llama la base de datos de emociones más grande del mundo, que consiste en más de diez millones de expresiones de personas de 87 países.[12] Trabajadores de *crowdsourcing*, la mayoría desde El Cairo, etiquetaron a mano su monumental colección de videos de personas exteriorizando sus emociones.[13] Muchas otras compañías han comprado los derechos de los productos de Affectiva para desarrollar de todo, desde aplicaciones para evaluar candidatos para puestos de trabajo hasta otras que analizan si los estudiantes están prestando atención en clases; todo esto capturando y analizando sus expresiones faciales y lenguaje corporal.[14]

Más allá del sector de las *startups*, todos los gigantes de la IA, como Amazon, Microsoft e IBM, tienen sus propios sistemas diseñados para detectar emociones y afectos. Microsoft ofrece una detección emocional en su Face API que asegura poder detectar lo que un individuo está sintiendo entre un rango de emociones que va de "ira, desprecio, asco, miedo, felicidad, neutralidad, tristeza y sorpresa", y afirma que "estas emociones son entendidas como transculturales y se comunican universalmente con expresiones faciales particulares".[15] De forma similar, la herramienta Rekognition de Amazon asegura que puede identificar "todas las siete emociones" y "medir cómo estas cosas pueden cambiar en el tiempo, una especie de línea temporal de las emociones de un actor".[16]

[8] Javier Sánchez-Monedero y Lina Dencik, "The Datafication of the Workplace", documento de trabajo, Data Justice Lab, Cardiff University, 9 de mayo de 2019, p. 48, disponible en línea: <https://datajusticeproject.net/wp-content/uploads/sites/30/2019/05/Report-The-datafication-of-the-workplace.pdf>, y Drew Harwell, "A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job", en *The Washington Post*, 7 de noviembre de 2019, disponible en línea: <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/>.

[9] Sam Byford, "Apple Buys Emotient, a Company That Uses AI to Read Emotions", en *The Verge*, 7 de enero de 2016, disponible en línea: <https://www.theverge.com/2016/1/7/10731232/apple-emotient-ai-startup-acquisition>.

[10] Phillip Molnar, Gary Robbins y David Pierson, "Cutting Edge: Apple's Purchase of Emotient Fuels Artificial Intelligence Boom in Silicon Valley", en *Los Angeles Times*, 17 de enero de 2016, disponible en línea: <https://www.latimes.com/business/technology/la-fi-cutting-edge-facial-recognition-20160117-story.html>.

[11] Rosalind Picard, "Affective Computing Group", en *MIT Media Lab*, disponible en línea: <https://affect.media.mit.edu/>.

[12] "Affectiva Human Perception AI Analyzes Complex Human States", en Affectiva, disponible en línea: <https://www.affectiva.com/>.

[13] Oscar Schwartz, *op. cit.*

[14] Véase, por ejemplo, Patricia Nilsson, *op. cit.*

[15] "Face: An AI Service That Analyzes Faces in Images", en Microsoft Azure, disponible en línea: <https://azure.microsoft.com/en-us/services/cognitive-services/face/>.

[16] "Amazon Rekognition Improves Face Analysis", en Amazon Web Services, 12 de agosto de 2019, disponible en línea: <https://aws.amazon.com/

Pero ¿cómo funcionan estas tecnologías? Los sistemas de reconocimiento emocional crecieron en los intersticios entre tecnologías de IA, las prioridades militares y las ciencias del comportamiento (la psicología en particular). Comparten un conjunto similar de cianotipos y suposiciones fundacionales: que hay un pequeño número de categorías emocionales distintas y universales, que revelamos esas emociones de forma involuntaria en nuestros rostros y que las máquinas las pueden detectar. Estos artículos de fe son tan aceptados en algunos campos que puede incluso parecer raro prestarles atención, y todavía más cuestionarlos; están tan integrados que han venido a constituir "el punto de vista común".[17] Pero si vemos cómo se llegó a taxonomizar las emociones (ordenadas y etiquetadas cuidadosamente), veremos que van surgiendo preguntas a cada paso. Y una de las figuras destacadas detrás de este método es la de Paul Ekman.

"El lector de rostros más famoso del mundo"

La investigación de Ekman comenzó con un afortunado encuentro con Silvan Tomkins, por entonces reconocido psicólogo con base en Princeton que había publicado en 1962 el primer volumen de su *magnum opus, Affect Imagery Consciousness.*[18]

---

about-aws/whats-new/2019/08/amazon-rekognition-improves-face-analysis/>, y "Amazon Rekognition - Video and Image - AWS", en Amazon Web Services, disponible en línea: <https://aws.amazon.com/rekognition/>.

[17] Lisa Feldman Barrett *et al., op. cit.,* p. 1.

[18] Eve Kosofsky Sedgwick, Adam Frank e Irving Alexander (eds.), *Shame and Its Sisters. A Silvan Tomkins Reader,* Durham, Duke University Press, 1995, p. 258.

El trabajo de Tomkins con las emociones tuvo una gran influencia en Ekman, que dedicó gran parte de su carrera a estudiar sus implicaciones. Un aspecto en particular cumplió un papel descomunal: la idea de que, si las emociones eran un conjunto innato de respuestas evolutivas, estas serían universales y, por lo tanto, reconocibles a través de las culturas. Este deseo de universalidad resulta muy relevante a la hora de entender por qué estas teorías se aplican ampliamente en los sistemas de reconocimiento emocional de la IA hoy en día: ofrecía un pequeño conjunto de principios que podían aplicarse en todas partes, una simplificación de la complejidad que era fácilmente replicable.

En la introducción a *Affect Imagery Consciousness,* Tomkins enmarcó su teoría de las emociones universales de base biológica como una que abordaba una crisis aguda de la soberanía humana. Desafiaba con esto el desarrollo del conductismo y el psicoanálisis, dos escuelas del pensamiento que, según él, trataban la conciencia como si fuera un mero subproducto de otras fuerzas que, a su vez, servían. Señalaba que la conciencia humana había "sido desafiada y reducida una y otra vez, primero por Copérnico", que desplazó al hombre desde el centro del universo, "después por Darwin", cuya teoría de la evolución destrozó por completo la idea de que los seres humanos habían sido creados a imagen de un Dios cristiano, "y sobre todo por Freud", que les quitó a la conciencia humana y a la razón su espacio como fuerzas impulsoras de nuestra motivación.[19] Tomkins continuaba su argumento con esta aseveración: "La paradoja del control máximo sobre la naturaleza, y el control mínimo sobre la naturaleza humana, deriva en parte del descuido de la función de la

[19] Silvan S. Tomkins, *Affect Imagery Consciousness. The Complete Edition,* Nueva York, Springer, 2008.

conciencia como un mecanismo de control".[20] Para decirlo de una manera más simple: *la conciencia nos dice muy poco respecto a por qué sentimos y actuamos como lo hacemos.* Esta es una afirmación clave para todo tipo de aplicaciones posteriores de la teoría de las emociones, ya que pone énfasis en la incapacidad de los humanos de reconocer tanto el sentimiento como la expresión de las emociones. Si como seres humanos somos incapaces de detectar realmente lo que sentimos, entonces ¿quizá los sistemas de IA pueden hacerlo por nosotros?

La teoría de las emociones de Tomkins fue su manera de abordar el problema de la motivación humana. Argumentó que la motivación estaba gobernada por dos sistemas: emociones e impulsos. Tomkins sostenía que los impulsos solían estar asociados estrechamente con las necesidades biológicas inmediatas, como el hambre y la sed,[21] y eran instrumentales: el dolor que causa el hambre se puede remediar con comida. Pero el sistema primario que gobierna la motivación humana y el comportamiento es el de las emociones, que incluye *sentimientos* positivos y negativos. Las emociones, que desempeñan el papel más importante en la motivación humana, amplifican las señales de los impulsos, pero son mucho más complejas. Por ejemplo, resulta difícil saber la razón precisa o las causas por las que llora un bebé, expresando la emoción del sufrimiento-angustia. El bebé puede estar "hambriento o [con] frío o mojado o adolorido o [llorando] por una fiebre".[22] De igual modo, hay un gran número de maneras en que se puede lidiar con este sentimiento emocional: "El llanto se puede detener alimentando el bebé,

[20] Silvan S. Tomkins, *op. cit.*

[21] Ruth Leys, *The Ascent of Affect. Genealogy and Critique,* Chicago, University of Chicago Press, 2017, p. 18.

[22] Silvan S. Tomkins, *op. cit.,* p. 23.

meciéndolo, subiendo la temperatura de la habitación, bajándola, quitando el alfiler del pañal, etcétera".[23]

Tomkins concluye lo siguiente:

> El precio que se paga por esta flexibilidad es el de la ambigüedad y el error. El individuo puede (o no) identificar correctamente la "causa" de su miedo o alegría, y puede (o no) aprender a reducir su miedo o mantener o recapturar su alegría. En este aspecto, el sistema de las emociones no es tan simple como un sistema de señales, como [sí] lo es el sistema de impulsos.[24]

Las emociones, como los impulsos, no son estrictamente instrumentales; tienen un grado más alto de independencia de los estímulos y los objetos, lo que quiere decir que, a menudo, podemos no saber por qué sentimos enojo, miedo, o felicidad.[25]

Toda esta ambigüedad puede sugerir que las complejidades de las emociones son imposibles de desenredar. ¿Cómo podemos saber cualquier cosa sobre un sistema en el que las conexiones entre la causa y el efecto, el estímulo y la respuesta, son tan tenues e inciertas? Tomkins propuso una respuesta: "Las emociones primarias [...] parecen están relacionadas innatamente de una forma individualizada con un sistema de órganos que es extraordinariamente visible". En otras palabras, el rostro.[26] Encontró preceden-

[23] *Ibid.*

[24] *Ibid.*

[25] Para Leys, esta "disociación radical entre sentimiento y cognición" es la principal razón de su atractivo para los teóricos de las humanidades, entre ellos Eve Kosofsky Sedgwick, quienes buscan revalorizar nuestras experiencias de error o confusión como nuevas formas de libertad. Ruth Leys, *op. cit.,* p. 35, y Eve Kosofsky Sedgwick, *Touching Feeling. Affect, Pedagogy, Performativity,* Durham, Duke University Press, 2003 [trad. esp.: *Tocar la fibra. Afecto, pedagogía, performatividad,* Madrid, Alpuerto, 2018].

[26] Silvan S. Tomkins, *op. cit.,* p. 204.

tes para este énfasis en la expresión facial en dos trabajos publicados en el siglo XIX: *La expresión de las emociones en el hombre y en los animales* (1872), de Charles Darwin, y un volumen desconocido escrito por el neurólogo francés Guillaume Benjamin Amand Duchenne de Boulogne, *Mécanisme de la physionomie humaine ou Analyse électro-physiologique de l'expression des passions applicable à la pratique des arts plastiques* (1862).[27]

Tomkins asumió que la exhibición facial de las emociones era un universal humano.

> Las emociones —creía— son conjuntos de respuestas musculares, vasculares y glandulares localizadas en el rostro y también distribuidas ampliamente por el cuerpo, lo que genera respuestas sensoriales. [...] Estos conjuntos organizados de respuestas son detonados en centros subcorticales donde se guardan "programas" específicos para cada emoción distinta;

este es uno de los primeros usos de una metáfora computacional para un sistema humano.[28]

Pero Tomkins reconoce que la *interpretación* de las manifestaciones emocionales depende de factores individuales, sociales y culturales. Admitió también que había muchos diferentes "dialectos" del lenguaje facial en distintas sociedades.[29] Es decir que hasta el progenitor de la investigación de las emociones plan-

[27] Silvan S. Tomkins, *op. cit.*, p. 206; Charles Darwin, *The Expression of the Emotions in Man and Animals*, ed. de Joe Cain y Sharon Messenger, Londres, Penguin, 2009 [trad. esp.: *La expresión de las emociones en el hombre y los animales*, Valencia, Sempere, s. f.], y Guillaume Benjamin Duchenne de Boulogne, *Mécanisme de la physionomie humaine ou Analyse électro-physiologique de l'expression des passions applicable à la pratique des arts plastiques*, París, Librairie J.-B. Baillière et Fils, 1876.

[28] Silvan S. Tomkins, *op. cit.*, p. 243, cit. en Ruth Leys, *op. cit.*, p. 32.

[29] *Ibid.*, p. 216.

teó la posibilidad de que el reconocimiento de los afectos y las emociones dependiera del contexto social y cultural. El conflicto potencial entre los dialectos culturales y un lenguaje universal basado en la biología tuvo enormes implicaciones para el estudio de las expresiones faciales y posteriores formas de reconocimiento emocional. Debido a que las expresiones faciales son culturalmente variables, usarlas para entrenar sistemas de aprendizaje automático supone inevitablemente mezclar todo tipo de contextos, señales y expectativas distintos.

A mediados de la década de 1960, la oportunidad golpeó a la puerta de Ekman. Se trataba de la Advanced Research Projects Agency (ARPA), una de las ramas de investigación del Departamento de Defensa. Al recordar este período, Ekman admitió: "No fue idea mía [la investigación de las emociones]. Se me pidió... me empujaron. Ni siquiera escribí la propuesta de investigación. La escribió por mí el hombre que me dio el dinero para hacerla".[30] En 1965, estaba investigando la expresión no verbal en ambientes clínicos y buscando fondos para desarrollar un programa de investigación en la Stanford University. Consiguió una entrevista en Washington con Lee Hough, que estaba a la cabeza de la división de Ciencias del Comportamiento de ARPA.[31] A Hough no le impresionó la descripción que Ekman hizo de su investigación, pero vio el potencial en el entendimiento de la comunicación no verbal transcultural.[32]

El único problema era que, según admitía el propio Ekman, él no sabía cómo hacer investigación transcultural: "Ni siquiera

[30] Paul Ekman, *Nonverbal Messages: Cracking the Code. My Life's Pursuit*, San Francisco, PEG, 2016, p. 45.

[31] Anna Tuschling, "The Age of Affective Computing", en Marie-Luise Angerer, Bernd Bösel y Michaela Ott (eds.), *Timing of Affect. Epistemologies, Aesthetics, Politics*, Zúrich, Diaphanes, 2014, p. 186.

[32] Paul Ekman, *Nonverbal Messages*, *op. cit.*, p. 45.

sabía cuáles eran los argumentos, la bibliografía o los métodos".[33] Así que, comprensiblemente, decidió dejar de buscar el financiamiento de ARPA. Pero Hough insistió y, de acuerdo con Ekman, "se sentó un día entero en mi oficina y escribió la propuesta que luego él mismo financió y me permitió hacer la investigación por la que soy conocido: la evidencia de la universalidad de ciertas expresiones faciales de emociones y las diferencias culturales en los gestos".[34] Obtuvo una inyección masiva de fondos de ARPA, cercana al millón de dólares; el equivalente a más de ocho millones de dólares en la actualidad.[35]

En ese momento, Ekman se preguntó por qué Hough parecía estar tan ansioso por financiar su investigación, a pesar de sus propias objeciones y su falta de experiencia. Resultó que Hough quería distribuir su dinero rápidamente para poder eludir las sospechas del senador Frank Church, que había descubierto a Hough usando la investigación en las ciencias sociales para tapar la adquisición de información en Chile que pudiera ser usada para derrocar al gobierno de izquierda del presidente Salvador Allende.[36] Ekman llegaría a concluir que era simplemente un tipo con suerte, alguien "capaz de hacer investigación en el extranjero que no lo metería [a Hough] en problemas".[37] ARPA sería la primera de una larga lista de agencias de defensa, inteligencia y policía que financiarían tanto la carrera de Ekman como, más ampliamente, el campo del reconocimiento emocional.

Con el apoyo de una gran beca a sus espaldas, Ekman empezó sus primeros estudios para probar la universalidad de las

[33] Paul Ekman, *Nonverbal Messages, op. cit.*, p. 46.
[34] *Ibid.*
[35] *Ibid.*
[36] *Ibid.*
[37] *Ibid.*

expresiones faciales. En general, estos estudios siguieron un diseño que sería replicado en los primeros laboratorios de IA. Ekman duplicó en gran parte los métodos de Tomkins, incluso usando las fotografías de este para hacer pruebas con sujetos tomados de Chile, Argentina, Brasil, Estados Unidos y Japón.[38] Confiaba en pedirles a los participantes de la investigación que simularan las expresiones de una emoción y, luego, las compararía con expresiones reunidas "en condiciones salvajes", esto es, fuera de un laboratorio.[39] Se les mostraban a los sujetos fotografías de expresiones faciales en pose, seleccionadas por los diseñadores, como ejemplares o expresivas de una emoción particularmente "pura" o intensa. Se les pedía entonces que eligieran entre esas categorías de emociones y etiquetaran la imagen. El análisis midió el grado en que las etiquetas elegidas por los sujetos se correlacionaban con las elegidas por los diseñadores.

Desde el principio, la metodología presentó problemas. El formato de respuesta de elección forzada de Ekman sería criticado más tarde por alertar a los sujetos de las conexiones que los diseñadores ya habían hecho entre expresiones faciales y emociones.[40] Además, el hecho de que estas emociones fueran falsificadas o actuadas levantaría preocupaciones significativas

[38] Paul Ekman y Erika L. Rosenberg (eds.), *What the Face Reveals. Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, Nueva York, Oxford University Press, 1997, p. 375.
[39] Silvan S. Tomkins y Robert McCarter, "What and Where Are the Primary Affects? Some Evidence for a Theory", en *Perceptual and Motor Skills*, vol. 18, núm. 1, 1964, pp. 119-158, disponible en línea: <https://doi.org/10.2466/pms.1964.18.1.119>.
[40] James A. Russell, "Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies", en *Psychological Bulletin*, vol. 115, núm. 1, 1994, p. 116, disponible en línea: <https://doi.org/10.1037/0033-2909.115.1.102>.

acerca de la validez de los resultados.[41] Ekman encontró algunos acuerdos transculturales usando este método, pero el antropólogo Ray Birdwhistell puso en duda sus hallazgos y sugirió que podían no reflejar estados emocionales innatos, si es que habían sido aprendidos culturalmente a través de la exposición a medios masivos como películas, televisión o revistas.[42] Fue esta controversia la que empujó a Ekman a dirigirse a Papúa Nueva Guinea, específicamente a estudiar indígenas en la región de las zonas altas. Pensó que, si las personas con poco contacto con la cultura y los medios occidentales coincidían con la forma en que había categorizado expresiones emocionales propuestas, esto le proporcionaría una fuerte evidencia de la universalidad de su esquema.

Después de regresar de su primer intento por estudiar a los fore en Papúa Nueva Guinea, Ekman ideó un enfoque alternativo para probar su teoría. Les mostró a sujetos de investigación en Estados Unidos una fotografía, luego les pidió que eligieran uno entre seis conceptos emocionales: felicidad, miedo, disgusto-desprecio, ira, sorpresa y tristeza.[43] Los resultados se acercaban lo suficiente a los de sujetos de otros países, con lo que Ekman creyó que estaba en condiciones de asegurar que "ciertos comportamientos faciales particulares están asociados universalmente con emociones particulares".[44]

[41] Ruth Leys, *op. cit.,* p. 93.

[42] Paul Ekman y Erika L. Rosenberg (eds.), *What the Face Reveals, op. cit.,* p. 377.

[43] Paul Ekman, E. Richard Sorenson y Wallace V. Friesen, "Pan-Cultural Elements in Facial Displays of Emotion", en *Science,* vol. 164, 1969, pp. 86 y 87, disponible en línea: <https://doi.org/10.1126/science.164.3875.86>.

[44] Paul Ekman y Wallace V. Friesen, "Constants across Cultures in the Face and Emotion", en *Journal of Personality and Social Psychology,* vol. 17, núm. 2, 1971, p. 128, disponible en línea: <https://doi.org/10.1037/h0030377>.

Emociones: de la fisonomía a la fotografía

La idea de que los estados interiores se pueden inferir de forma confiable a partir de señales exteriores surge en parte de la historia de la fisonomía, que fue premisa en el estudio de los rasgos faciales de las personas para encontrar indicaciones sobre su carácter. En el mundo griego antiguo, Aristóteles había creído que "es posible juzgar el carácter de los hombres a partir de su aspecto físico [...] pues se ha asumido que el cuerpo y el alma se ven afectados juntos".[45] Los griegos también usaban la fisonomía como una forma temprana de clasificación racial, aplicada al "género del hombre mismo, dividiéndolo en razas, en la medida en que difirieran en apariencia y en personalidad (por ejemplo, egipcios, tracios y escitas)".[46] Presuponían una conexión entre el cuerpo y el alma que justificaba una lectura del carácter interior de una persona basada en su apariencia externa.

La fisonomía en la cultura occidental llegó a un punto álgido durante los siglos XVIII y XIX, cuando se la veía como parte de las ciencias anatómicas. Una figura central de esta tradición era el pastor suizo Johann Kaspar Lavater, quien escribió *Essays on Physiognomy; for the Promotion of Knowledge and the Love of Mankind* [Ensayos sobre la fisonomía para la promoción del conocimiento y el amor a la humanidad], publicado originalmente en alemán en 1789.[47] Lavater tomó los métodos de la fisonomía

[45] Aristóteles, *The Categories. On Interpretation,* trad. de Harold Percy Cooke y Hugh Tredennick, Cambridge, Harvard University Press, 1938, 70b, 8-13, 527 [trad. esp.: *Categorías,* Buenos Aires, Colihue, 2009].

[46] *Ibid.,* 805a, 27-30, 87.

[47] Sería difícil exagerar la influencia de esta obra, que desde entonces ha caído en descrédito: en 1810 ya tenía dieciséis ediciones en alemán y veinte en inglés. John Graham, "Lavater's Physiognomy in England", en *Journal of*

y los mezcló con los últimos conocimientos científicos. Intentó crear una comparación más "objetiva" de los rostros usando siluetas en lugar de grabados de artistas, ya que estas eran más mecánicas y fijaban la posición de cada rostro en una forma familiar (la del perfil), lo que permitía un punto de vista comparativo.[48] Creía que la estructura ósea era una conexión subyacente entre la apariencia física y el tipo de carácter. Mientras que las expresiones faciales eran pasajeras, los cráneos ofrecían un material más sólido para las inferencias fisonómicas.[49] Como vimos en el capítulo anterior, la medición de cráneos se utilizó para apoyar el nacionalismo, el racismo y la xenofobia emergentes. A lo largo del siglo XIX, elaboraron este infame trabajo frenólogos como Franz Joseph Gall y Johann Gaspar Spurzheim, así como el trabajo de Cesare Lombroso en la criminología científica... todo lo cual conduciría a los tipos de clasificaciones inferenciales que recurren en los sistemas de IA contemporáneos.

Pero fue el neurólogo francés Duchenne, descripto por Ekman como un "observador maravillosamente dotado", quien codificó el uso de la fotografía y otros medios técnicos en el estudio de los rostros humanos.[50] En *Mécanisme de la physionomie humaine,* Duchenne sentó importantes bases tanto para Darwin como para Ekman, conectando viejas ideas de la fisonomía y la frenología con investigaciones más modernas de la

the History of Ideas, vol. 22, núm. 4, 1961, p. 561, disponible en línea: <https://doi.org/10.2307/2708032>.

[48] Richard T. Gray, *About Face. German Physiognomic Thought from Lavater to Auschwitz,* Detroit, Wayne State University Press, 2004, p. 342.

[49] Jean-Jacques Courtine y Claudine Haroche, *Histoire du visage. Exprimer et taire ses émotions (du XVIe siècle au début du XIXe siècle),* París, Payot & Rivages, 2007, p. 132.

[50] Paul Ekman, "Duchenne and Facial Expression of Emotion", en Guillaume Benjamin Duchenne de Boulogne, *The Mechanism of Human Facial Expression,* ed. y trad. de R. A. Cuthbertson, Cambridge, Cambridge University Press, 1990, pp. 270-284.

fisiología y la psicología. Remplazó vagas afirmaciones sobre el carácter con una investigación más limitada sobre la expresión y los estados mentales o emocionales interiores.[51]

Duchenne trabajaba en París, en el asilo Salpêtrière, que alojaba hasta cinco mil personas con un amplio rango de diagnósticos de enfermedades mentales y condiciones neurológicas. Algunos se convertirían en sujetos de sus angustiosos experimentos, parte de una larga tradición de experimentación médica y tecnológica con aquellos más vulnerables que no pueden negarse.[52] Duchenne, que era poco conocido en la comunidad científica, desarrolló técnicas de descargas eléctricas para estimular movimientos musculares aislados en los rostros de las personas. Su meta era construir un entendimiento anatómico y fisiológico más completo del rostro. Usó estos métodos para unir la nueva ciencia psicológica con el estudio mucho más antiguo de los signos fisonómicos, o pasiones.[53] Confió en las últimas técnicas fotográficas, como el procesamiento de colodión, que permitía tiempos de exposición mucho más cortos, lo que le permitió congelar en imágenes movimientos musculares y expresiones faciales fugaces.[54]

Incluso en estas etapas muy tempranas, los rostros nunca fueron expresiones humanas naturales o que ocurrieran en contextos sociales, sino *simulaciones* producidas por la aplicación

[51] Guillaume Benjamin Duchenne de Boulogne, *op. cit.*

[52] François Clarac, Jean Massion y Allan M. Smith, "Duchenne, Charcot and Babinski, Three Neurologists of La Salpetrière Hospital, and Their Contribution to Concepts of the Central Organization of Motor Synergy", en *Journal of Physiology - Paris,* vol. 103, núm. 6, 2009, pp. 362 y 363, disponible en línea: <https://doi.org/10.1016/j.jphysparis.2009.09.001>.

[53] François Delaporte, *Anatomy of the Passions,* trad. de Susan Emanuel, Stanford, Stanford University Press, 2008, p. 33 [trad. esp.: *Anatomía de las pasiones,* Barranquilla, Uninorte, 2007].

[54] *Ibid.,* pp. 48-51.

bruta de electricidad en los músculos. Aun así, Duchenne creía que el uso de la fotografía y otros sistemas técnicos transformaría el laxo asunto de la representación en algo objetivo y probatorio, más pertinente para el estudio científico.[55] En su introducción a *La expresión de emociones en el hombre y en los animales,* Darwin elogió las "magníficas fotografías" de Duchenne e incluyó reproducciones en su propio trabajo.[56] Puesto que las emociones son ocurrencias temporales, incluso fugaces, la fotografía ofrecía la posibilidad de fijar, comparar y categorizar su aparición en el rostro. Sin embargo, las imágenes verdaderas de Duchenne estaban muy fabricadas.

Ekman seguiría a Duchenne al poner a la fotografía en el centro de su práctica experimental.[57] Creía que la fotografía en cámara lenta era esencial para su método, porque muchas expresiones faciales operan en los límites de la percepción humana. Buscaba encontrar las llamadas microexpresiones, movimientos musculares minúsculos en la cara. La duración de las microexpresiones, según él, "es tan corta que, a menos que se use la proyección en cámara lenta, se quedan en el umbral del reconocimiento".[58] En años posteriores, Ekman también insistiría en que cualquiera podía llegar a aprender a reconocer microexpresiones, sin un entrenamiento especial ni el uso de capturas en cámara lenta, en alrededor de una hora.[59] Pero si estas expresiones son dema-

[55] Lorraine Daston y Peter Galison, *Objectivity,* Nueva York, Zone Books, 2010.

[56] Charles Darwin, *op. cit.,* pp. 12 y 307.

[57] Ruth Leys, *op. cit.,* p. 85, y James A. Russell, "Is There Universal Recognition of Emotion from Facial Expression? A Review of the Cross-Cultural Studies", en *Psychological Bulletin,* vol. 115, núm. 1, 1994, p. 114, disponible en línea: <https://doi.org/10.1037/0033-2909.115.1.102>.

[58] Paul Ekman y Wallace V. Friesen, "Nonverbal Leakage and Clues to Deception", en *Psychiatry,* vol. 31, núm. 1, 1969, p. 93.

[59] Mark Williams Pontin, "Lie Detection", en *MIT Technology Review,* 21 de abril de 2009, disponible en línea: <https://www.technologyreview.com/s/413133/lie-detection/>.

Placas de Guillaume Benjamin Duchenne (de Boulogne), Mécanisme de la physionomie humaine. *Cortesía de la Biblioteca Nacional de Medicina de Estados Unidos.*

siado rápidas para que los humanos puedan reconocerlas, ¿cómo van a poder entenderlas?[60]

Uno de los planes más ambiciosos de Ekman al principio de su investigación era codificar un sistema para detectar y analizar expresiones faciales.[61] En 1971, copublicó la descripción de lo que llamó Facial Action Scoring Technique (FAST) [Técnica de puntuación de acciones faciales]. Apoyándose en fotografías para las que habían posado, el método usaba seis tipos emocionales básicos, derivados principalmente de las intuiciones de Ekman.[62] Pero FAST pronto se encontró con problemas cuando otros científicos fueron capaces de producir expresiones faciales que no estaban incluidas en su tipología.[63] Así que Ekman decidió basar su nueva herramienta de medición en la musculatura facial, remontándose a los estudios originales de electrochoque de Duchenne. Ekman identificó aproximadamente cuarenta contracciones musculares distintas en el rostro, y llamó a los componentes básicos de cada expresión facial unidad de acción.[64] Después de algunas pruebas y validaciones Ekman y Wallace Friesen publicaron el Facial Action Coding System (FACS) [Sistema de codificación de la acción facial] en 1978; las ediciones actualizadas continúan siendo ampliamente utilizadas.[65] FACS era extremadamente laborioso para ser usado como una herramienta de medición. Ekman dijo que le tomaba entre setenta y cinco y ochenta horas capacitar a los usuarios en la metodología FACS, y una hora obtener un minuto de metraje facial.[66]

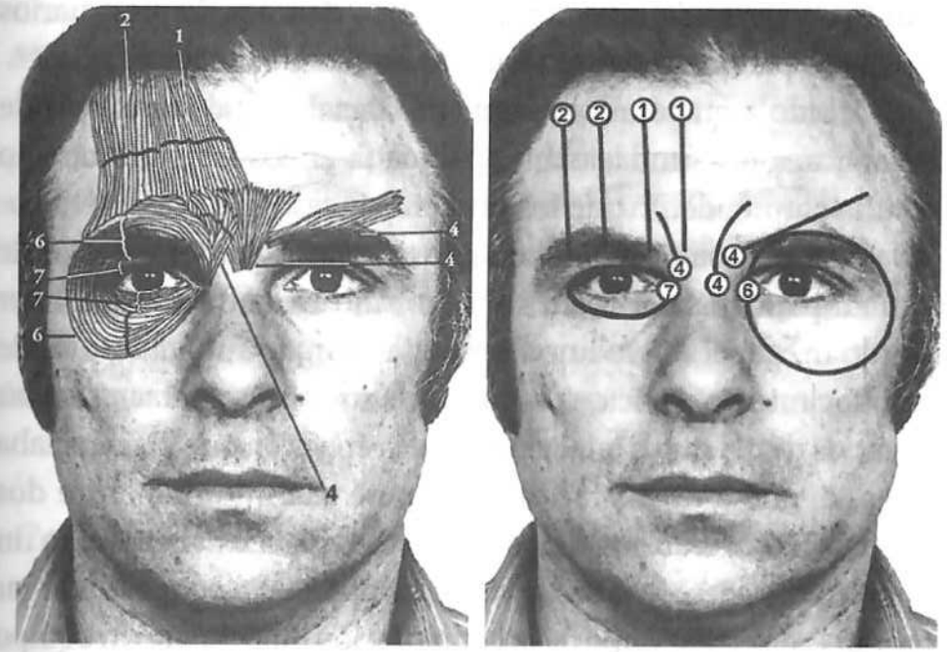
[60] Paul Ekman y Wallace V. Friesen, "Nonverbal Leakage...", *op. cit.*, p. 94. En una nota al pie, los autores explican: "Nuestra propia investigación y la evidencia de la neurofisiología de la percepción visual sugieren fuertemente que se pueden percibir microexpresiones que son tan cortas como un cuadro de película (1/50 de segundo). El hecho de que estas microexpresiones no se vean habitualmente debe depender de que estén incrustadas en otras expresiones que distraen la atención, su poca frecuencia o algún hábito perceptual aprendido de ignorar las expresiones faciales rápidas".

[61] Paul Ekman, E. Richard Sorenson y Wallace V. Friesen, "Pan-Cultural Elements in Facial Displays of Emotion", en *Science,* vol. 164, 1969, p. 87, disponible en línea: <https://doi.org/10.1126/science.164.3875.86>.

[62] Paul Ekman, Wallace V. Friesen y Silvan Tomkins, "Facial Affect Scoring Technique: A First Validity Study", en *Semiotica*, vol. 3, 1971, p. 40.

[63] Paul Ekman, *Nonverbal Messages, op. cit.*, p. 97.

[64] *Ibid.*, p. 102.



Elementos del Sistema de codificación de la acción facial.
Fuente: Paul Ekman y Wallace V. Friesen.

En una conferencia a principios de los años ochenta, Ekman escuchó una presentación que sugería una solución a las intensas demandas laborales de FACS: el uso de computadoras para automatizar las mediciones. Aunque en sus memorias Ekman no menciona al investigador que dio la charla, sí señala que el sistema se llamaba Wizard y que había sido desarrollado en la Brunel Uni-

[65] Paul Ekman y Erika L. Rosenberg (eds.), *What the Face Reveals, op. cit.*

[66] Paul Ekman, *Nonverbal Messages, op. cit.*, p. 105.

versity de Londres.[67] Seguramente, se trata del sistema de reconocimiento de objetos de aprendizaje automático de Igor Aleksander, WISARD, uno de los primeros de su tipo, que usaba redes neuronales en una época en que este método estaba pasado de moda.[68] Algunas fuentes reportan que WISARD estaba entrenado con "una base de datos de conocidos *hooligans* de fútbol", anticipando el difundido uso contemporáneo de fotos de prontuarios policiales para entrenar tecnologías de reconocimiento facial.[69]

Debido a que el reconocimiento facial surgió como una de las aplicaciones fundacionales para la IA en los años sesenta, no resulta sorprendente que los primeros investigadores que trabajaban en este campo hicieran causa común con el método de Ekman para analizar rostros.[70] El mismo Ekman asegura haber tenido un papel activo impulsando las formas automatizadas de reconocimiento emocional a través de sus viejos contactos en las agencias de defensa e inteligencia, de la época en que lo financiaba ARPA. Ayudó a organizar una competencia informal entre dos equipos trabajando con datos FACS, y esto parece haber tenido un impacto duradero. Ambos equipos han pasado a ocupar un lugar prominente en el campo de la computación emocional. Un equipo estaba compuesto por Terry Sejnowski y su estudiante Marian Bartlett, que se convertiría en una figura importante de la informática de reconocimiento emocional y la científica principal de Emotient, adquirida por Apple en 2016.[71] El segundo equipo, con

[67] Paul Ekman, *Nonverbal Messages, op. cit.*, p. 169.

[68] *Ibid.*, p. 106, e Igor Aleksander (ed.), *Artificial Vision for Robots,* Boston, Springer US, 1983.

[69] "Magic from Invention", en Brunel University London, disponible en línea: <https://www.brunel.ac.uk/research/Brunel-Innovations/Magic-from invention>.

[70] Woodrow Wilson Bledsoe, "The Model Method in Facial Recognition", reporte técnico, PRI 15, Palo Alto, Panoramic Research, 1964.

[71] Phillip Molnar, Gary Robbins y David Pierson, *op. cit.*

sede en Pittsburgh, era dirigido por el psicólogo Jeffrey Cohn de la University of Pittsburgh y el eminente investigador de visión artificial Takeo Kanade de Carnegie Mellon.[72] Estas dos figuras se dedicaron al reconocimiento emocional a largo plazo y desarrollaron el muy conocido conjunto de datos de expresión emocional Cohn-Kanade (CK) y sus descendientes.

El sistema FACS de Ekman proporcionó dos cosas esenciales a las aplicaciones posteriores de aprendizaje automático: un conjunto de etiquetas estable, discreto y finito, que los humanos podían usar para categorizar fotografías de rostros, así como un sistema para producir mediciones. Prometía desplazar, del ámbito de los artistas y novelistas, el difícil trabajo de representar las vidas interiores para ponerlo bajo la protección de una rúbrica racional, cognoscible y mensurable, adecuada para los laboratorios, corporaciones y gobiernos.

CAPTURAR EL SENTIMIENTO: EL ARTIFICIO DETRÁS DE LA INTERPRETACIÓN DE LAS EMOCIONES

A medida que empezaba a tomar forma el uso de computadoras para el reconocimiento emocional, los investigadores admitieron que necesitaban una recolección de imágenes estandarizadas para experimentar. Un informe de la National Science Foundation (NSF) de 1992, escrito en colaboración con Ekman, recomendaba que "una base de datos multimedia, de fácil acceso, compartida por la diversa comunidad de la investigación facial, sería un importante recurso para resolver y ampliar algunos temas rela-

[72] Takeo Kanade, *Computer Recognition of Human Faces,* Basel, Birkhäuser Boston, 2013.

cionados con el entendimiento facial".[73] En menos de un año, el Departamento de Defensa comenzaría a financiar el programa Face Recognition Technology (FERET) para recolectar fotografías faciales, como vimos en el capítulo III. Para el final de la década, los investigadores de aprendizaje automático habían comenzado a compilar, etiquetar y hacer públicos los conjuntos de datos que impulsan la mayor parte de la investigación de aprendizaje automático hoy en día.

Las pautas de FACS de Ekman formaron directamente el conjunto de datos de CK.[74] Siguiendo con la tradición de Ekman de posar para las expresiones faciales, "a los sujetos se les instruyó por medio de un experimentador a realizar una serie de 23 demostraciones faciales", que los expertos de FACS luego codificaron, proporcionando etiquetas para los datos. El conjunto de datos de CK permitió a los laboratorios comparar sus resultados y su progreso a medida que construían nuevos sistemas de reconocimiento de expresiones.

Otros laboratorios y compañías trabajaban en proyectos paralelos, creando decenas de bases de datos. Por ejemplo, los investigadores de un laboratorio en Suecia crearon el Karolinska Directed Emotional Faces. Esta base de datos se compone de imágenes de individuos que posan con expresiones emocionales que se corresponden con las categorías de Ekman.[75] Hacen que sus

[73] Takeo Kanade, *Computer Recognition of Human Faces, op. cit.*, p. 16.

[74] Takeo Kanade, Jeffrey F. Cohn y Yingli Tian, "Comprehensive Database for Facial Expression Analysis", en *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition,* 2000, p. 46, disponible en línea: <https://doi.org/10.1109/AFGR.2000.840611>.

[75] Véanse Takeo Kanade, Jeffrey F. Cohn y Yingli Tian, *op. cit.;* Michael Lyons *et al.*, "Coding Facial Expressions with Gabor Wavelets", en *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition,* 1998, pp. 200-205, disponible en línea: <https://doi.org/10.1109/AFGR.1998.670949>, y Elen Goeleven *et al.*, "The Karolinska Directed Emotional Faces: A Validation Study", en *Cognition and Emotion,* vol. 22, núm. 6,

rostros adopten formas acordes a seis estados emocionales básicos. Al mirar estos conjuntos de datos de entrenamiento cuesta no sorprenderse por lo extremos que son: *¡Sorpresa increíble! ¡Felicidad abundante! ¡Miedo paralizante!* Estos sujetos están literalmente creando emociones legibles por computadora.

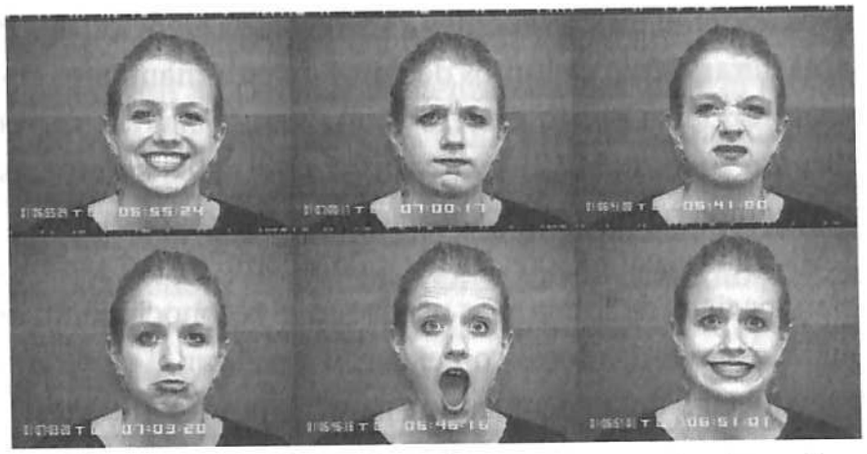
Mientras el campo crecía en escala y complejidad, también lo hacían los tipos de fotografías usadas en el reconocimiento emocional. Los investigadores comenzaron a utilizar el sistema FACS para etiquetar datos generados no a partir de poses con expresiones, sino de expresiones faciales espontáneas, a veces recopiladas fuera del laboratorio. Por ejemplo, una década después del inmensamente exitoso lanzamiento del conjunto de datos CK, un grupo de investigadores lanzó una segunda generación, el conjunto de datos Cohn-Kanade Extendido (CK+),[76] que incluía el rango típico de poses con expresiones, pero también empezó a incluir las llamadas expresiones sin pose o espontáneas, tomadas de videos en los que los sujetos hacían expresiones faciales espontáneas.

Para 2009, Affectiva surgió del MIT Media Lab con el objetivo de capturar "expresiones faciales naturalistas y espontáneas" en escenarios de la vida real.[77] La compañía recolectó datos permitiendo a los usuarios participar en un sistema que grabaría sus rostros usando una cámara web mientras miraban una serie de comerciales. Estas imágenes serían luego etiquetadas usando un

2008, pp. 1094-1018, disponible en línea: <https://doi.org/10.1080/02699930701626582>.

[76] Patrick Lucey *et al.*, "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression", en *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops,* pp. 94-101, disponible en línea: <https://doi.org/10.1109/CVPRW.2010.5543262>.

[77] Daniel McDuff *et al.*, "Affectiva-MIT Facial Expression Dataset (AM-FED): Naturalistic and Spontaneous Facial Expressions Collected 'In-the-Wild'", en *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops,* pp. 881-888, disponible en línea: <https://doi.org/10.1109/CVPRW.2013.130>.



Expresiones faciales del conjunto de datos Cohn-Kanade: alegría, ira, disgusto, tristeza, sorpresa, miedo. Imágenes para las que se ha posado de Takeo Kanade et al., Yearbook of Physical Anthropology, *2000. © Cohn & Kanade.*

software personalizado por codificadores entrenados con el FACS.[78] Pero aquí encontramos otro problema de circularidad: FACS fue desarrollado a partir del cuantioso archivo de Ekman de fotografías en que se posa;[79] incluso cuando las imágenes son recolectadas en escenarios naturalistas, por lo general se clasifican de acuerdo con un esquema derivado de imágenes para las que se posa.

El trabajo de Ekman influyó profunda y ampliamente en casi todo, desde el *software* de detección de mentiras hasta la visión artificial. *The New York Times* lo describió como "el lector de rostros más famoso del mundo" y *Time* lo nombró una de las cien personas más influyentes del mundo. Eventualmente, llegaría a ser asesor de clientes tan dispares como el Dalai Lama, el Federal Bureau of Investigation (FBI), la Intelligence Agency (CIA), el Servicio Secreto e, incluso, el estudio de animación Pixar, que

[78] Daniel McDuff *et al., op. cit.*

[79] Paul Ekman y Wallace V. Friesen, *Facial Action Coding System (FACS). A Technique for the Measurement of Facial Action,* Palo Alto, Consulting Psychologists Press, 1978.

quería crear representaciones más realistas para las caras de sus dibujos animados.[80] Sus ideas se volvieron parte de la cultura popular, incluidas en *best-sellers* como *Blink: inteligencia intuitiva,* de Malcolm Gladwell, y un drama de televisión, *Miénteme,* en el que Ekman trabajó como asesor del personaje principal, aparentemente basado vagamente en él.[81]

Sus negocios también prosperaron: Ekman vendió técnicas de detección de engaños a agencias de seguridad como la Transportation Security Administration (TSA), que las usó para desarrollar el programa Screening of Passengers by Observation Techniques (SPOT). SPOT se usó para monitorear las expresiones faciales de los viajeros aéreos en los años que siguieron a los ataques del 11 de septiembre, en un intento por detectar terroristas "automáticamente". El sistema usa un conjunto de 94 criterios que, supuestamente, son señales de estrés, miedo o engaño. Pero estar buscando estas respuestas significaba que algunos grupos estaban inmediatamente en desventaja. Cualquiera que estuviera estresado, que se sintiera incómodo al ser interrogado o que hubiera tenido experiencias negativas con la policía o los controles fronterizos podía obtener una puntuación más alta. Esto produjo su propia forma de discriminación racial. El programa SPOT ha sido criticado por la Government Accountability Office (GAO) y por grupos que abogan por los derechos indivi-

[80] Judy Foreman, "A Conversation with: Paul Ekman; The 43 Facial Muscles That Reveal Even the Most Fleeting Emotions", en *The New York Times,* 5 de agosto de 2003, disponible en línea: <https://www.nytimes.com/2003/08/05/health/conversation-with-paul-ekman-43-facial-muscles-that-reveal-even-most-fleeting.html>; Jill Bolte Taylor, "The 2009 Time 100", en *Time,* 30 de abril de 2009, disponible en línea: <http://content.time.com/time/specials/packages/article/0,28804,1894410_1893209_1893475,00.html>, y Paul Ekman Group, disponible en línea: <https://www.paulekman.com/>.

[81] Sharon Weinberger, "Airport Security: Intent to Deceive?", en *Nature,* vol. 465, 2010, p. 413, disponible en línea: <https://doi.org/10.1038/465412a>.

duales, por su falta de metodología científica y por no producir éxitos claros, a pesar de su precio de 900 millones de dólares.[82]

Las críticas a las teorías de Ekman

Cuanto más crecía la fama de Ekman, también lo hacía el escepticismo hacia su trabajo, con críticas que provenían de distintos campos. Una de las primeras críticas fue la antropóloga cultural Margaret Mead, que debatió con Ekman sobre la cuestión de la universalidad de las emociones a fines de la década de 1960, lo que resultó en intercambios feroces no solo entre ellos dos, sino entre otros antropólogos que criticaban las ideas de Ekman sobre una universalidad absoluta.[83] A Mead no la convencía que Ekman creyera en determinantes biológicos universales del comportamiento y no considerara factores culturales.[84] En particular, Ekman tendía a reducir las emociones a un binario demasiado simplificado y mutuamente excluyente: o las emociones eran universales o no lo eran. Los críticos como Mead señalaban que había otros matices.[85] Ella adoptó una posición intermedia y destacó que no había ninguna contradicción inherente entre "la

[82] Ashley Halsey III, "House Member Questions $900 Million TSA 'SPOT' Screening Program", en *The Washington Post,* 14 de noviembre de 2013, disponible en línea: <https://www.washingtonpost.com/local/trafficandcommuting/house-member-questions-900-million-tsa-spot-screening-program/2013/11/14/ad194cfe-4d5c-11e3-be6b-d3d28122e6d4_story.html>.

[83] Paul Ekman, "A Life's Pursuit", en Thomas A. Sebeok y Jean Umiker-Sebeok (eds.), *The Semiotic Web '86: An International Yearbook,* Berlín, Mouton de Gruyter, 1987, pp. 4-46, y *Nonverbal Messages, op. cit.,* pp. 79-81.

[84] Margaret Mead, "Review of *Darwin and Facial Expression. A Century of Research in Review,* edited by Paul Ekman", en *Journal of Communication,* vol. 25, núm. 1, 1975, p. 209, disponible en línea: <https://doi.org/10.1111/j.1460-2466.1975.tb00574.x>.

[85] Silvan S. Tomkins, *Affect Imagery Consciousness, op. cit.,* p. 216.

posibilidad de que los seres humanos puedan compartir un núcleo de comportamientos innatos [...] y la idea de que las expresiones emocionales puedan, *al mismo tiempo,* estar muy condicionadas por factores culturales".[86]

Otros científicos de distintos campos se sumaron al coro en las siguientes décadas. En años más recientes los psicólogos James Russell y José Miguel Fernández-Dols han demostrado que los aspectos más básicos de la ciencia siguen estando sin resolver: "Las preguntas más fundamentales, como por ejemplo si acaso 'las expresiones faciales de emociones' expresan de hecho emociones, permanecen sujetas a grandes controversias".[87] Las científicas sociales Maria Gendron y Lisa Feldman Barrett han señalado los peligros específicos de usar las teorías de Ekman en la industria de la IA, porque la detección automática de expresiones faciales no indica con fiabilidad un estado mental interno.[88] Como observa

[86] Margaret Mead, *op. cit.,* p. 212. Véase también Alan Fridlund, "A Behavioral Ecology View of Facial Displays, 25 Years Later", en *Emotion Researcher,* agosto de 2015, disponible en línea: <https://emotionresearcher.com/the-behavioral-ecology-view-of-facial-displays-25-years-later/>. Ekman terminaría aceptando muchos de los postulados de Mead. Véanse Paul Ekman, "Argument for Basic Emotions", en *Cognition and Emotion,* vol. 6, núm. 3-4, 1992, pp. 169-200; *Emotions Revealed, op. cit.,* y "What Scientists Who Study Emotion Agree About", en *Perspectives on Psychological Science,* vol. 11, núm. 1, 2016, pp. 81-88, disponible en línea: <https://doi.org/10.1177/1745 691615596992>. A Ekman tampoco le faltaban defensores. Véanse Alan Cowen *et al.,* "Mapping the Passions: Toward a High-Dimensional Taxonomy of Emotional Experience and Expression", en *Psychological Science in the Public Interest,* vol. 20, núm. 1, 2019, pp. 61-90, disponible en línea: <https://doi.org/10.1177/1529100619850176>, y Hillary Anger Elfenbein y Nalini Ambady, "On the Universality and Cultural Specificity of Emotion Recognition: A Meta-Analysis", en *Psychological Bulletin,* vol. 128, núm. 2, 2002, pp. 203-235, disponible en línea: <https://doi.org/10.1037/0033-2909.128.2.203>.

[87] José Miguel Fernández-Dols y James A. Russell (eds.), *The Science of Facial Expression,* Nueva York, Oxford University Press, 2017, vol. 1, p. 4.

[88] Maria Gendron y Lisa Feldman Barrett, "Facing the Past", en José Miguel Fernández-Dols y James A. Russell (eds.), *op. cit.,* vol. 1, p. 30.

Barrett, "las compañías pueden decir lo que quieran, pero los datos son claros. Pueden detectar un ceño fruncido, pero eso no es lo mismo que detectar enojo".[89]

Todavía más problemático resulta que, en el campo de estudio de las emociones, no haya consenso entre los investigadores respecto a lo que es en realidad una emoción. Qué son, cómo se formulan dentro nuestro y se expresan, sus funciones fisiológicas y neurobiológicas, su relación con los estímulos, incluso cómo definirlas: todo esto sigue total y obstinadamente lejos de ser resuelto.[90]

Quizá la principal crítica de la teoría de las emociones de Ekman sea la historiadora científica Ruth Leys. En *The Ascent of Affect* desglosa minuciosamente

> las implicancias de las suposiciones fundamentalmente fisiognómicas que subyacen en el trabajo de Ekman [...] en concreto, la idea de que se puede mantener una distinción estricta entre expresiones emocionales auténticas y artificiales, basándose en las diferencias entre las caras que hacemos cuando estamos solos y las que hacemos cuando estamos con otros.[91]

[89] James Vincent, "AI 'Emotion Recognition' Can't Be Trusted", en *The Verge,* 25 de julio de 2019, disponible en línea: <https://www.theverge.com/2019/7/25/8929793/emotion-recognition-analysis-ai-machine-learning-facial-expression-review>. Los académicos que estudian discapacidades también han notado que las suposiciones sobre cómo funcionan la biología y los cuerpos pueden suscitar preocupaciones en torno al sesgo, especialmente cuando se automatizan a través de la tecnología. Véase Meredith Whittaker *et al.,* "Disability, Bias, and AI", en AI Now Institute, noviembre de 2019, disponible en línea: <https://ainowinstitute.org/disabilitybiasai-2019.pdf>.

[90] Carroll E. Izard, "The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation", en *Emotion Review,* vol. 2, núm. 4, 2010, pp. 363-370, disponible en línea: <https://doi.org/10.1177/1754073910374661>.

[91] Ruth Leys, *op. cit.,* p. 22.

Leys ve una circularidad fundamental en el método de Ekman. En primer lugar, las fotografías en las que se posa o se simula que usó expresaban un conjunto de estados emocionales básicos, "ya libres de influencias culturales".[92] Después, estas fotografías se usaban para obtener etiquetas de distintas poblaciones para demostrar la universalidad de las expresiones faciales. Leys señala un problema serio: Ekman asumió que "las expresiones faciales en las fotografías que usó en sus experimentos estaban libres de máculas culturales porque habían sido reconocidas a nivel universal. Al mismo tiempo, sugirió que esas expresiones faciales eran reconocidas universalmente porque estaban libres de máculas culturales".[93] El método es en esencia recursivo.[94]

Otros problemas se volvieron evidentes cuando las ideas de Ekman empezaron a utilizarse en sistemas técnicos. Como hemos visto, muchos conjuntos de datos que forman la base de este campo están basados en actores *simulando* estados emocionales, actuando para la cámara. Esto significa que los sistemas de IA están entrenados para reconocer expresiones fingidas de sentimientos. Aunque los sistemas de IA aseguran tener acceso a la verdad fundamental sobre los estados naturales interiores, están entrenados con materiales que, inevitablemente, son construidos. Incluso en las imágenes que capturan a personas respondiendo a comerciales o películas, esas personas son conscientes de que están siendo observadas, lo que puede alterar sus reacciones.

La dificultad para automatizar la conexión entre los movimientos faciales y las categorías emocionales básicas nos lleva a un cuestionamiento más amplio: si las emociones pueden, de hecho, agruparse adecuadamente en pequeños números de cate-

[92] *Ibid.,* p. 92.
[93] *Ibid.,* p. 94.
[94] *Ibid.*

gorías discretas.[95] Este punto de vista se remonta a Tomkins, quien argumentaba que "cada tipo de emoción puede identificarse mediante una respuesta distintiva más o menos única dentro del cuerpo".[96] Pero hay muy poca evidencia consistente que apunte a esto. Los psicólogos han llevado a cabo muchas revisiones de la evidencia publicada, sin encontrar las asociaciones entre las respuestas medibles y los estados emocionales, cuya existencia esta evidencia asume.[97] Finalmente, está el obstinado asunto de que las expresiones faciales a veces no indican precisamente nuestros estados interiores, como podrá confirmar quienquiera que alguna vez haya sonreído sin sentirse feliz.[98]

Ninguno de estos importantes cuestionamientos a la base de las aseveraciones de Ekman ha impedido que su trabajo tenga un lugar privilegiado en las aplicaciones de IA actuales. Cientos de artículos citan la visión de Ekman sobre las expresiones faciales interpretables, como si se tratara de un hecho indiscutible, a pesar de las décadas de controversia científica. Muy pocos científicos informáticos han siquiera reconocido esta literatura de incertidumbres. El investigador de computación afectiva Arvid Kappas, por ejemplo, menciona directamente una falta de consenso científico básico:

> Sabemos demasiado poco sobre los complejos moduladores sociales de la actividad facial, y posiblemente de otras actividades expresivas en tales situaciones, como para poder medir de manera confiable los estados emocionales a partir de la conducta expresiva. *Este no es un problema de ingeniería que puede resolverse con un mejor algoritmo.*[99]

[95] Lisa Feldman Barrett, "Are Emotions Natural Kinds?", en *Perspectives on Psychological Science,* vol. 1, núm. 1, 2006, p. 28, disponible en línea: <https://doi.org/10.1111/j.1745-6916.2006.00003.x>.

[96] *Ibid.,* p. 30.

[97] Véase, por ejemplo, Lisa Feldman Barrett *et al., op. cit.*

[98] *Ibid.,* p. 40.

A diferencia de muchos en este campo, que confían ciegamente en el reconocimiento emocional, Kappas cuestiona por completo la creencia de que es una buena idea que las computadoras intenten detectar las emociones.[100]

Entre más tiempo pasan investigadores de otros campos examinando el trabajo de Ekman, más fuerte se vuelve la evidencia en su contra. En 2019, Lisa Feldman Barrett dirigió un equipo de investigación que realizó una amplia revisión de la bibliografía sobre la inferencia de emociones a partir de expresiones faciales. Concluyeron tajantemente que las expresiones faciales están lejos de ser indiscutibles, y que "no son 'huellas digitales' o muestras de diagnóstico" que señalan con fiabilidad estados emocionales, mucho menos a través de culturas y contextos. Basándose en toda la evidencia actual, el equipo señaló que "no es posible inferir con seguridad la felicidad a partir de una sonrisa, la rabia de un ceño fruncido o la tristeza de una mueca, como intenta hacer gran parte de la tecnología actual al aplicar lo que equivocadamente se cree que son hechos científicos".[101]

El equipo de Barrett criticaba a las compañías de IA que aseguraban ser capaces de automatizar la inferencia de las emociones:

> Las compañías tecnológicas, por ejemplo, están gastando millones de dólares de investigación para construir dispositivos que

[99] Arvid Kappas, "Smile When You Read This, Whether You Like It or Not: Conceptual Challenges to Affect Detection", en *IEEE Transactions on Affective Computing,* vol. 1, núm. 1, 2010, p. 39, disponible en línea: <https://doi.org/10.1109/T-AFFC.2010.6> (el énfasis me pertenece).

[100] *Ibid.,* p. 40.

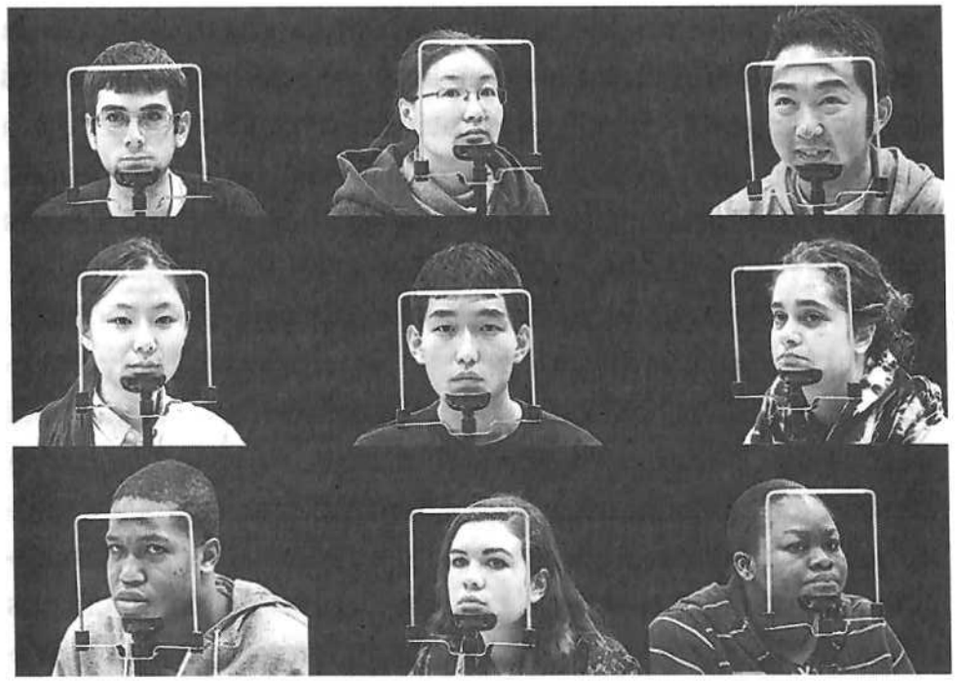
[101] Lisa Feldman Barrett *et al., op. cit.,* p. 46.

> lean emociones de los rostros, tomando erróneamente la visión común como si fuera algo que tiene un fuerte apoyo científico. [...] De hecho, nuestra revisión de la evidencia científica indica que se conoce muy poco acerca de cómo y por qué ciertos movimientos faciales expresan instancias emocionales, particularmente con un nivel de detalle suficiente como para que semejantes conclusiones se utilicen en programas importantes del mundo real.[102]

¿Por qué, teniendo tantas críticas, ha permanecido el método de "leer emociones" de los rostros? Al analizar la historia de estas ideas, nos damos cuenta de que el financiamiento de la investigación militar, las prioridades policiales y el lucro han delineado esta área. Desde los años sesenta, impulsados por importantes fondos del Departamento de Defensa, se han desarrollado muchos sistemas que son cada vez más precisos en la medición de *movimientos* en los rostros. Una vez que surgió la teoría de que era posible evaluar estados internos midiendo movimientos faciales, y que se desarrolló la tecnología para medirlos, la gente adoptó voluntariamente la premisa subyacente. La teoría calzaba con lo que las herramientas podían hacer. Las teorías de Ekman parecían ideales para el campo emergente de la visión artificial porque podían automatizarse a escala.

Existen poderosas inversiones institucionales y corporativas detrás de la validez de las teorías y metodologías de Ekman. Reconocer que las emociones no se pueden clasificar fácilmente, o que no se pueden detectar de manera confiable en expresiones faciales, podría socavar a una industria en expansión. En el campo de la IA, se suele citar a Ekman como si el tema se hubiera resuelto, antes de proceder directamente a desafíos de la ingeniería.

[102] Lisa Feldman Barrett *et al., op. cit.,* pp. 47 y 48.



Conjunto de datos Columbia Gaze, en Brian A. Smith et al., ACM Symposium on User Interface Software and Technology (UIST), *octubre de 2013, pp. 271-280. Cortesía de Brian A. Smith.*

Los asuntos más complejos de contexto, acondicionamiento, relacionalidad y factores culturales son difíciles de conciliar con los métodos disciplinarios actuales de la informática, o con las ambiciones del sector tecnológico comercial. Por lo tanto, las categorías emocionales básicas de Ekman se volvieron estándar. Métodos más sutiles, como el intermedio de Mead, se pasaron en gran medida por alto. El enfoque se ha puesto en aumentar las tasas de precisión de los sistemas de IA, en lugar de abordar las mayores preguntas respecto a las muchas maneras en que experimentamos, mostramos y ocultamos emociones y cómo interpretamos las expresiones faciales de los demás.

Como escribe Barrett: "Muchos de los modelos más influyentes de nuestra ciencia asumen que las emociones son categorías biológicas impuestas por la naturaleza, de modo que las catego-

rías emocionales son *reconocidas,* en vez de construidas, por la mente humana".[103] Los sistemas de IA para detectar emociones se basan en esta idea. El reconocimiento puede ser el marco equivocado por completo cuando se piensa en las emociones, porque el reconocimiento asume que las categorías emocionales están determinadas, en lugar de ser emergentes y relacionales.

La política de los rostros

En vez de intentar construir más sistemas que agrupen expresiones en categorías legibles por computadora, deberíamos preguntarnos por los orígenes de esas categorías, así como por sus consecuencias sociales y políticas. Actualmente, se usan herramientas de reconocimiento emocional para ataques políticos. Por ejemplo, un blog conservador aseguró haber creado un "sistema de polígrafo virtual" para evaluar videos de la congresista Ilhan Abdullahi Omar.[104] Mediante análisis de rostro y voz de Rekognition de Amazon, xrVision Sentinel AI y Watson de IBM, el bloguero aseguraba que el puntaje analítico de mentiras de Omar superaba la "media de verdad" y que sus registros de estrés, desprecio y nerviosismo eran altos. Varios medios de comunicación conservadores reportaron la noticia, alegando que Omar era una "mentirosa patológica" y una amenaza para la seguridad de la nación.[105]

[103] Lisa Feldman Barrett *et al., op. cit.,* p. 47 (el énfasis me pertenece).

[104] Yaacov Apelbaum, "One Thousand and One Nights and Ilhan Omar's Biographical Engineering", en *The Illustrated Primer* (blog), 13 de agosto de 2019, disponible en línea: <https://apelbaum.wordpress.com/2019/08/13/one-thousand-and-one-nights-and-ilhan-omars-biographical-engineering/>.

[105] Véase, por ejemplo, Joe Hoft, "Facial, Speech and Virtual Polygraph Analysis Shows Ilhan Omar Exhibits Many Indications of a Compulsive Fibber!!!", en *The Gateway Pundit,* 21 de julio de 2019, disponible en línea: <https://www.thegatewaypundit.com/2019/07/facial-speech-and-virtual-polygraph-analysis-shows-ilhan-omar-exhibits-many-indications-of-a-compulsive-fibber/>.

Es sabido que estos sistemas señalan las afectaciones del habla de las mujeres de manera diferente del modo en que señalan las de los hombres, en especial las de las mujeres negras. Como vimos en el capítulo III, la construcción de un "promedio" a partir de datos de entrenamiento no representativos desde el principio resulta epistemológicamente sospechoso, con sesgos claramente raciales. Un estudio llevado a cabo en la University of Maryland demostró que algunos *softwares* de reconocimiento facial interpretan los rostros negros como si tuvieran mayores emociones negativas que los blancos, en particular registrándolos como más enojados y desdeñosos, incluso llegando a medir el grado de sus sonrisas.[106]

Este es el peligro de las herramientas de reconocimiento emocional. Como hemos visto, nos devuelven a un pasado frenológico, en el que se hicieron aseveraciones falsas que fueron no solo permitidas, sino desplegadas para apoyar los sistemas de poder existentes. Las décadas de controversias científicas sobre la inferencia de emociones distintas a partir de rostros humanos subraya un punto central: el modelo de reconocimiento "de talla única" no es la metáfora correcta para identificar estados emocionales. Las emociones son completas y se desarrollan y cambian en relación con nuestras familias, amigos, culturas e historias, y todos los diferentes contextos que viven fuera del marco de la IA. En muchos casos, los sistemas de detección de emociones no hacen lo que aseguran. En lugar de medir directamente los estados mentales interiores de las personas, simplemente optimizan estadísticamente correlaciones de ciertas características físicas entre

[106] Lauren Rhue, "Racial Influence on Automated Perceptions of Emotions", en *Social Science Research Network,* 9 de noviembre de 2018, disponible en línea: <https://dx.doi.org/10.2139/ssrn.3281765>.

imágenes faciales. Las bases científicas de la detección de emociones automatizada están puestas en duda y, sin embargo, una nueva generación de herramientas emocionales ya está haciendo inferencias a través de una gama cada vez más amplia de contextos de alto riesgo, de la vigilancia policial a la contratación.

A pesar de que la evidencia ahora apunta a la poca fiabilidad de la detección emocional, las compañías siguen buscando nuevas fuentes de minería para la imaginería facial, compitiendo por liderar la participación en el mercado de un sector que promete miles de millones en beneficios. La revisión sistemática de Barrett de la investigación detrás de la inferencia de emociones a partir de los rostros de las personas concluye con una nota condenatoria:

> De manera más general, las compañías de tecnología puede que se estén haciendo una pregunta que en su esencia está equivocada. Los esfuerzos por simplemente "leer" los estados internos de las personas a partir tan solo de un análisis de sus movimientos faciales, sin considerar varios aspectos del contexto, son en el mejor de los casos insuficientes, y en el peor carecen completamente de validez, sin importar cuán sofisticados sean los algoritmos computacionales. [...] Resulta prematuro utilizar esta tecnología para sacar conclusiones sobre lo que la gente siente basándose en sus movimientos faciales.[107]

Hasta que no resistamos el deseo de automatizar el reconocimiento emocional, seguiremos corriendo el riesgo de juzgar de manera injusta a postulantes a trabajos porque sus microexpresiones no calzan con las de otros empleados, de que algunos estudiantes reciban peores notas que sus pares porque sus caras

[107] Lisa Feldman Barrett *et al.*, "Emotional Expressions Reconsidered", *op. cit.*, p. 48.

indicaban falta de entusiasmo y de que se detenga a algunos clientes porque un sistema de IA los marcó como posibles ladrones únicamente basándose en sus características faciales.[108] Estas son las personas que pagarán los costos de tener sistemas que no solo son imperfectos técnicamente hablando, sino que además están basados en metodologías cuestionables.

Las áreas de la vida en que estos sistemas operan se están expandiendo con la misma velocidad con que los laboratorios y las empresas pueden crear nuevos mercados para ellos. Y, sin embargo, todos dependen de un entendimiento muy limitado de las emociones, derivado del conjunto inicial de Ekman de ira, felicidad, sorpresa, asco, tristeza y miedo, que se supone que han de representar el universo infinito de sentimientos y expresiones humanos a través del espacio y el tiempo. Esto nos lleva de vuelta a las profundas limitaciones a la hora de capturar las complejidades del mundo en un solo esquema clasificatorio. Nos devuelve al mismo problema que hemos visto repetidamente: el deseo de simplificar en extremo lo que resulta obstinadamente complejo para computarse y empaquetarse fácilmente para el mercado. Los sistemas de IA están buscando extraer las experiencias mutables, privadas y divergentes de nuestros yoes corporales, pero el resultado es una caricatura que no consigue capturar los matices de la experiencia emocional en el mundo.

[108] Véanse, por ejemplo, Neil Connor, "Chinese School Uses Facial Recognition to Monitor Student Attention in Class", en *Telegraph,* 17 de mayo de 2018, disponible en línea: <https://www.telegraph.co.uk/news/2018/05/17/chinese-school-uses-facial-recognition-monitor-student-attention/>, y Lisa Du y Ayaka Maki, "AI Cameras That Can Spot Shoplifters Even before They Steal", en *Bloomberg,* 4 de marzo de 2019, disponible en línea: <https://www.bloomberg.com/news/articles/2019-03-04/the-ai-cameras-that-can-spot-shoplifters-even-before-they-steal>.

Traducción de
FRANCISCO DIAZ KLAASSEN

KATE CRAWFORD

ATLAS DE INTELIGENCIA ARTIFICIAL

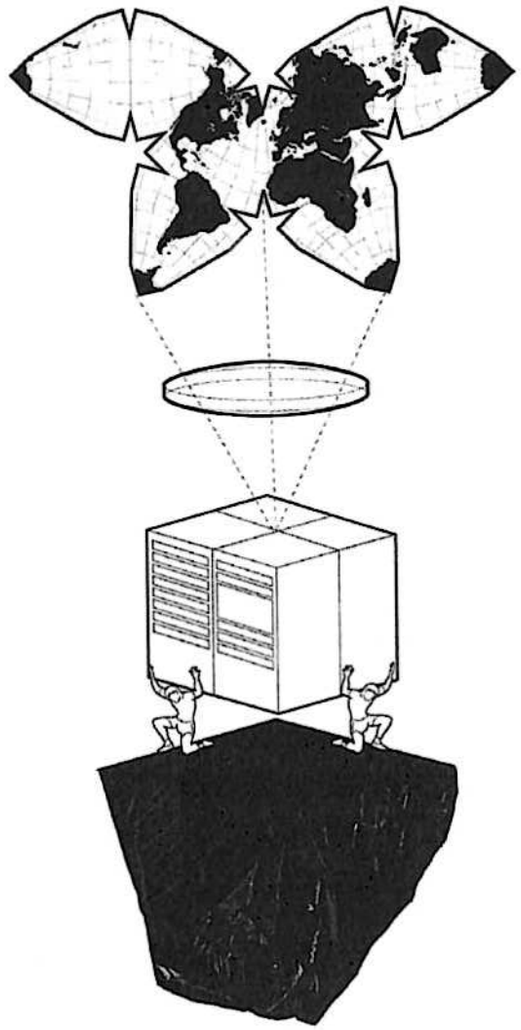
Poder, política y costos planetarios

FONDO DE CULTURA ECONÓMICA
MÉXICO - ARGENTINA - BRASIL - CHILE - COLOMBIA - ECUADOR - ESPAÑA
ESTADOS UNIDOS DE AMÉRICA - GUATEMALA - PERÚ - VENEZUELA

de AI Now Institute, que han dado forma a este libro: *AI Now 2019 Report,* en AI Now Institute, 2019 (con Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Amba Kak, Elizabeth Kaziunas, Varoon Mathur, Erin McElroy, Andrea Nill Sánchez, Deborah Raji, Joy Lisi Rankin, Rashida Richardson, Jason Schultz, Sarah Myers West y Meredith Whittaker); "Discriminating Systems: Gender, Race and Power in AI", AI Now Institute, 2019 (con Sarah Myers West y Meredith Whittaker); en *AI Now Report 2018,* en AI Now Institute, 2018 (con Meredith Whittaker, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Myers West, Rashida Richardson, Jason Schultz y Oscar Schwartz); "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability", AI Now Institute, 2018 (con Dillon Reisman, Jason Schultz y Meredith Whittaker); en *AI Now 2017 Report,* en AI Now Institute, 2017 (con Alex Campolo, Madelyn Sanfilippo, y Meredith Whittaker), y *AI Now 2016 Report,* NYU Information Law Institute, 2016 (con Madeleine Clare Elish, Solon Barocas, Aaron Plasek, Kadija Ferryman y Meredith Whittaker).

Finalmente, este libro no existiría sin las siguientes personas: Trevor Paglen, una verdadera brújula, desde las exploraciones en el desierto hasta las investigaciones arqueológicas; Vladan Joler, un amigo en la elaboración de mapas, cuyos diseños iluminan este libro y mi pensamiento; Laura Poitras, que me dio el valor; Karen Murphy, por su ojo de diseñadora; Adrian Hobbes y Edwina Throsby, por ayudarme a apagar incendios; Bo Daley, quien mejoró todo, y mi familia, Margaret, James, Judith, Claudia, Cliff y Hilary. Un agradecimiento eterno para Jason y Elliott, mis cartógrafos favoritos.

Introducción



El caballo más inteligente del mundo

A fines del siglo XIX, Europa se encontraba embelesada por un caballo llamado Hans. "Clever Hans"* era una verdadera maravilla: podía resolver problemas matemáticos, decir la hora, identificar días en el calendario, diferenciar tonos musicales y deletrear palabras y frases. La gente acudía en masa a ver al semental alemán golpeteando con el casco las respuestas a problemas complejos y dando consistentemente las respuestas correctas. "¿Cuánto es dos más tres?" De manera diligente, Hans chocaba un casco contra el suelo cinco veces. "¿Qué día de la semana es?" El caballo entonces se acercaba a un tablero especialmente diseñado y deletreaba la respuesta correcta usando el mismo método. Hans llegó incluso a dominar preguntas más complejas: "Estoy pensando en un número. Si le quito nueve, me quedan tres. ¿De qué número se trata?". Para 1904, Clever Hans ya era famoso en el mundo entero, y *The New York Times*

* "Hans el astuto", en castellano. A pesar de que el apelativo original era alemán ("Hans der Kluge"), mundialmente se lo conoció con la versión en inglés. [N. del T.]

promovía sus habilidades con el siguiente titular: "El maravilloso caballo de Berlín. ¡Solo le falta hablar!".[1]

El entrenador de Hans, Wilhelm von Osten –un profesor de matemáticas retirado–, había estado fascinado por la inteligencia animal durante mucho tiempo. Intentó sin éxito enseñarles a gatitos y crías de oso los números cardinales, y solo triunfó cuando empezó a trabajar con su propio caballo. Primero, le enseñó a contar sujetándole la pata, mostrándole un número y luego golpeteando con el casco el número correcto de veces. Pronto Hans fue capaz de realizar sumas simples. A continuación, Von Osten incorporó un pizarrón que tenía escrito el abecedario, de modo que el caballo pudiera hacer chocar un casco sobre un número para cada letra. Después de dos años de entrenamiento, Von Osten estaba asombrado por la capacidad del animal para comprender conceptos intelectuales avanzados. Así que se lo llevó de gira para probarle al mundo que los animales podían razonar. Hans se viralizó durante la *Belle Époque.*

Pero había muchos escépticos, y el consejo educativo alemán creó una comisión de investigación para comprobar la validez de las declaraciones científicas hechas por Von Osten. La Comisión Hans estaba dirigida por el psicólogo y filósofo Carl Stumpf y su asistente Oskar Pfungst, e incluía a un director de circo, un profesor retirado, un zoólogo, un veterinario y un oficial de caballería. Sin embargo, después de que interrogaran largamente a Hans, tanto con su entrenador presente como sin él, el caballo mantuvo su registro de respuestas correctas, y la comisión no fue capaz de encontrar evidencia alguna de fraude. Como escribiría más tarde Pfungst, Hans actuó frente a "miles de espectadores, aficionados a los caballos, entrenadores de calibre, y ninguno de ellos, durante el transcurso de observaciones que duraron muchos meses, descubrió algún tipo de señal secreta regular" entre el interrogador y el caballo.[2]

[1] Edward T. Heyn, "Berlin's Wonderful Horse; He Can Do Almost Everything but Talk - How He Was Taught", en *The New York Times,* 4 de septiembre de 1904, disponible en línea: <https://timesmachine.nytimes.com/timesmachine/1904/09/04/101396572.pdf>.

Wilhelm von Osten y Clever Hans.

La comisión descubrió que los métodos enseñados a Hans se asemejaban más a "la enseñanza en la educación básica" que al entrenamiento de animales, y eran "dignos de estudiarse científicamente".[3] Pero Stumpf y Pfungst todavía tenían algunas dudas. Un hallazgo en particular los preocupaba: cuando el mismo interrogador no sabía la respuesta a la pregunta que estaba for-

[2] Oskar Pfungst, *Clever Hans (The Horse of Mr. von Osten). A Contribution to Experimental Animal and Human Psychology,* trad. de Carl L. Rahn, Nueva York, Henry Holt, 1911, s. d.

[3] "'Clever Hans' Again: Expert Commission Decides that the Horse Actually Reasons", en *The New York Times,* 2 de octubre de 1904, disponible en línea: <https://timesmachine.nytimes.com/timesmachine/1904/10/02/120289067.pdf>.

mulando o cuando estaba muy lejos de Hans, este rara vez acertaba. Esto los llevó a considerar que quizás alguna señal involuntaria le estuviera dando las respuestas a Hans.

Como Pfungst describió en su libro de 1911, no se equivocaban: la postura del interrogador, así como su respiración y expresión facial, cambiaban sutilmente cuando los golpeteos de Hans alcanzaban la respuesta correcta, lo que lo hacía parar justo en ese momento.[4] Pfungst probó esta hipótesis más tarde en sujetos humanos y confirmó sus resultados. Para él, la parte deslumbrante de este descubrimiento era que, por lo general, los interrogadores no sabían que le estaban dando pistas al caballo. La solución al enigma de Clever Hans, escribió Pfungst, eran las indicaciones inconscientes de los interrogadores.[5] El caballo estaba entrenado para producir los resultados que su dueño quería ver, pero el público se sintió estafado porque esta no era la inteligencia extraordinaria que se habían imaginado.

La historia de Clever Hans es cautivante por muchos motivos: la relación entre el deseo, la ilusión y la acción; el negocio del espectáculo; la manera en que antropomorfizamos lo no humano; cómo emergen los sesgos y la política de la inteligencia. Hans inspiró un término en psicología que se usa para un tipo especial de trampa conceptual, el efecto Clever Hans, también llamado sesgo de expectativa, que describe la influencia que tienen las pistas involuntarias que los investigadores dan a los sujetos de investigación. La relación entre Hans y Von Osten tiene que ver con los complejos mecanismos a través de los cuales los sesgos se abren camino en los sistemas y cómo la gente termina enredada con los fenómenos que estudia. En la actualidad, la historia de Hans se usa en aprendizaje automático como un recordatorio

[4] Oskar Pfungst, *op. cit.*
[5] *Ibid.*

admonitorio de que no siempre se puede estar seguro de lo que ha aprendido un modelo a partir de los datos que se le han dado.[6] Incluso un sistema que parece funcionar de forma espectacular durante las prácticas puede hacer predicciones terribles cuando se le presentan datos nuevos en el mundo real.

Esto nos lleva a la pregunta central de este libro: ¿cómo se "hace" la inteligencia y con qué trampas nos podemos encontrar a partir de este proceso? A primera vista, la historia de Clever Hans es la historia de un hombre que construyó inteligencia entrenando a un caballo para que siguiera pistas y emulara la cognición humana. Pero, en otro nivel, podemos observar que la práctica de crear inteligencia se extendió considerablemente más allá de esta situación. El empeño requirió la validación de múltiples instituciones, incluida la academia, la educación, la ciencia, el público y las fuerzas armadas. Luego, está el mercado en el que se insertaron Von Osten y su notable caballo: inversiones tanto emocionales como económicas que estaban detrás de las giras, los artículos periodísticos, las charlas. Autoridades burocráticas congregadas para sopesar y analizar las habilidades del caballo. Una constelación de intereses financieros, culturales y científicos que cumplieron un rol en la construcción de la inteligencia de Hans y que tenían un interés tangible en que esta fuera realmente notable.

Aquí podemos ver dos mitologías distintas. El primer mito es que los sistemas no humanos (sean computadoras o caballos) son análogos a la mente humana. Esta perspectiva presupone que, con el entrenamiento adecuado o los recursos suficientes, una inteligencia parecida a la de un ser humano se puede crear de cero sin tener en consideración las maneras fundamentales

[6] Sebastian Lapuschkin *et al.*, "Unmasking Clever Hans Predictors and Assessing What Machines Really Learn", en *Nature Communications,* vol. 10, núm. 1, 2019, pp. 1-8, disponible en línea: <https://doi.org/10.1038/s41467-019-08987-4>.

en que las personas se encarnan, se relacionan y se ubican dentro de contextos más amplios. El segundo mito es que la inteligencia es algo que existe de forma independiente, como algo natural y separado de las fuerzas sociales, culturales, históricas y políticas. Este concepto de inteligencia ha causado un daño enorme durante siglos y se ha usado para justificar relaciones de dominación desde la esclavitud hasta la eugenesia.[7]

Estas mitologías prevalecen particularmente en el campo de la inteligencia artificial (IA), donde la creencia de que las máquinas pueden formalizar y reproducir la inteligencia humana ha sido axiomática desde mediados del siglo XX. Así como se consideraba la inteligencia de Hans equiparable a la de un ser humano, cuidadosamente fomentada como la de un niño durante su educación básica, en muchas ocasiones se han descripto los sistemas de IA como formas simples de inteligencia pero parecidas a la humana. En 1950, Alan Turing predijo que "para el fin de siglo el uso de palabras y opiniones generales razonadas habrá cambiado tanto que uno podrá hablar de máquinas pensantes sin esperar a que lo contradigan".[8] En 1958, el matemático John von Neumann aseguró que el sistema nervioso de los seres humanos es, "a primera vista, digital".[9] En cierta ocasión, ante la

[7] Véase el trabajo de la filósofa Val Plumwood sobre los dualismos inteligencia-estupidez, emocional-racional y maestro-esclavo. Val Plumwood, "The Politics of Reason: Towards a Feminist Logic", en *Australasian Journal of Philosophy,* vol. 71, núm. 4, 1993, pp. 436-462, disponible en línea: <https://doi.org/10.1080/00048409312345432>.

[8] Alan M. Turing, "Computing Machinery and Intelligence", en *Mind,* 1° de octubre de 1950, pp. 433-460, disponible en línea: <https://doi.org/10.1093/mind/LIX.236.433> [trad. esp.: "Maquinaria computacional e inteligencia", trad. de Cristóbal Fuentes Barassi para la Universidad de Chile, Santiago de Chile, 2010, disponible en línea: <http://xamanek.izt.uam.mx/map/cursos/Turing-Pensar.pdf>].

[9] John von Neumann, *The Computer and the Brain,* New Haven, Yale University Press, 1958, p. 44 [trad. esp.: *El ordenador y el cerebro,* Barcelona, Bosch,

pregunta de si las máquinas podían pensar, el profesor del Massachusetts Institute of Technology (MIT) Marvin Minsky respondió: "Por supuesto que las máquinas pueden pensar; nosotros podemos pensar y somos 'máquinas hechas de carne'".[10] Pero no todo el mundo estaba tan convencido. Uno de los precursores de la invención de la IA, Joseph Weizenbaum, creador de ELIZA, el primer programa conversacional, creía que la idea de que los seres humanos son meros sistemas procesadores de información involucraba una noción demasiado simplista de la inteligencia, y que esa noción estaba detrás de la "perversa fantasía de proporciones" que suponía pensar que los científicos de la IA llegarían a crear una máquina que aprendiera "al igual que un niño".[11]

Esta ha sido una de las discusiones centrales en la historia de la IA. En 1961, el MIT fue sede de una serie de importantes conferencias con el título de "Administración de empresas y la computadora del futuro". El programa incluía la participación de una serie estelar de científicos informáticos, entre los que se encontraban Grace Hopper, Joseph Carl Robnett Licklider,

1980]. Este método fue muy criticado por Hubert L. Dreyfus en *What Computers Can't Do. A Critique of Artificial Reason,* Nueva York, Harper and Row, 1972.

[10] Véase Joseph Weizenbaum, "On the Impact of the Computer on Society: How Does One Insult a Machine?", en *Science,* vol. 176, núm. 4035, 1972, p. 612. Tras su muerte, Minsky se vio implicado en serias acusaciones relacionadas con el pedófilo y violador convicto Jeffrey Epstein. Minsky era uno de los muchos científicos que se reunieron con Epstein y lo visitaron en su isla de retiro, donde se obligaba a niñas a tener sexo con los miembros de su camarilla. Como ha observado la académica Meredith Broussard, esto formaba parte de una cultura de exclusión más amplia, que se volvió endémica en la IA: "Por muy creativos que fueran Minsky y su séquito, también solidificaron la cultura de la tecnología como un club de amigos multimillonarios. Las matemáticas, la física y las otras ciencias 'duras' nunca han sido hospitalarias con las mujeres y la gente de color; la tecnología siguió este ejemplo". Véase Meredith Broussard, *Artificial Unintelligence. How Computers Misunderstand the World,* Cambridge, MIT Press, 2018, p. 174.

[11] Joseph Weizenbaum, *Computer Power and Human Reason. From Judgment to Calculation,* San Francisco, W. H. Freeman, 1976, pp. 202 y 203.

Marvin Minsky, Allen Newell, Herbert Simon y Norbert Wiener, para discutir los rápidos avances que se estaban produciendo en la computación digital. En su osada conclusión, John McCarthy argumentó que las diferencias entre las tareas humanas y las de las máquinas eran ilusorias. Se trataba simplemente de que las máquinas requerirían más tiempo para formalizar y resolver algunas de las tareas humanas más complicadas.[12]

Pero el profesor de filosofía Hubert Dreyfus contraargumentó, preocupado por el hecho de que los ingenieros en su conjunto "ni siquiera consideraran la posibilidad de que el cerebro pudiera procesar información de una manera completamente diferente a una computadora".[13] En una obra posterior, *What Computers Can't Do,* Dreyfus señala que la inteligencia y la pericia humanas dependen en gran medida de muchos procesos, inconscientes y subconscientes, mientras que las computadoras requieren que todos los procesos y los datos sean explícitos y estén formalizados.[14] Como resultado, los aspectos menos formales de la inteligencia tienen que ser abstraídos, eliminados y aproximados, lo que vuelve a las computadoras incapaces de procesar información sobre situaciones como lo harían los seres humanos.

Mucho ha cambiado en la IA desde la década de 1960, incluido un giro de los sistemas simbólicos a la más reciente ola hiperbólica sobre las técnicas de aprendizaje automático. En varios sentidos, las primeras discusiones sobre lo que puede hacer la IA han sido olvidadas y el escepticismo se ha desvanecido. Desde mediados de la década de 2000, la IA se ha expandido rápidamente, como campo académico y como industria. Hoy en día, un pequeño número de compañías tecnológicas poderosas hace uso de sistemas de IA a escala planetaria, y sus sistemas son aclamados, una vez más, como similares o incluso superiores a la inteligencia humana.

[12] Martin Greenberger (ed.), *Management and the Computer of the Future,* Nueva York, John Wiley and Sons, 1962, p. 315.

[13] Hubert L. Dreyfus, *Alchemy and Artificial Intelligence,* Santa Monica, RAND, 1965, s. d.

[14] Hubert L. Dreyfus, *What Computers Can't Do, op. cit.*

Sin embargo, la historia de Clever Hans también nos recuerda cuán estrechamente consideramos o reconocemos la inteligencia. A Hans se le enseñó a imitar tareas dentro de un rango muy acotado: sumar, restar y deletrear. Esto refleja una perspectiva limitada de lo que los caballos o los seres humanos pueden hacer. En términos de comunicación entre especies, Hans llevaba a cabo hazañas notables, además de actuar frente al público y tener una considerable paciencia; sin embargo, ninguna de estas acciones se reconocía como inteligencia. En palabras de la autora e ingeniera Ellen Ullman, esta creencia de que la mente es como una computadora, y viceversa, ha "infectado décadas de pensamiento en las ciencias de la computación y cognitivas", y se volvió una especie de pecado original en el campo.[15] Se trata de la ideología del dualismo cartesiano en la IA, en la que esta se limita a ser entendida como una inteligencia incorpórea, liberada de cualquier relación con el mundo material.

¿Qué es la inteligencia artificial? Ni artificial ni inteligente

Hagamos la pregunta, en apariencia simple, de qué es la IA. Si le preguntas a alguien en la calle, puede que mencione a Siri de Apple, la nube de Amazon, los autos de Tesla o el algoritmo de búsqueda de Google. Si les preguntas a los expertos en aprendizaje profundo, puede que te den una respuesta técnica acerca

[15] Ellen Ullman, *Life in Code. A Personal History of Technology,* Nueva York, MCD, 2017, pp. 136 y 137.

de cómo las redes neuronales se organizan en docenas de capas que reciben datos etiquetados a los que se les asignan pesos y niveles de capacidad para, de esta manera, poder clasificarlos de modos que aún no podemos explicar del todo.[16] En 1978, mientras discutía sobre los sistemas expertos, el profesor Donald Michie describió la IA como el refinamiento del conocimiento, donde "se puede producir una confiabilidad y competencia en la codificación que supera con creces el nivel más alto que un ser humano experto y sin ayuda haya podido, o quizás incluso pueda algún día, alcanzar".[17] En uno de los manuales más populares sobre el tema, Stuart Russell y Peter Norvig señalan que la IA es el intento por entender y construir entidades inteligentes. "La inteligencia se ocupa principalmente de la acción racional —aseguran—. De manera ideal, un agente inteligente toma la mejor decisión posible en una situación."[18]

Cada una de estas formas de definirla cumple con un cometido y establece un marco de referencia para entenderla, medirla, valorarla y gobernarla. Cuando a la IA la definen marcas de consumo para la infraestructura empresarial, entonces son el *marketing* y la publicidad las que han predeterminado el horizonte. Cuando los sistemas de IA se ven más confiables o racionales que cualquier experto humano, capaces de realizar la "mejor acción posible", entonces se está sugiriendo que debería

[16] Véase, como uno de muchos ejemplos, Tomaso Poggio *et al.*, "Why and When Can Deep –but not Shallow– Networks Avoid the Curse of Dimensionality: A Review", en *International Journal of Automation and Computing*, vol. 14, núm. 5, 2017, pp. 503-519, disponible en línea: <https://link.springer.com/article/10.1007/s11633-017-1054-2>.

[17] Cit. en Karamjit S. Gill, *Artificial Intelligence for Society*, Nueva York, John Wiley and Sons, 1986, p. 3.

[18] Stuart J. Russell y Peter Norvig, *Artificial Intelligence. A Modern Approach*, 3ª ed., Upper Saddle River, Pearson, 2010, p. 30 [trad. esp.: *Inteligencia artificial. Un enfoque moderno*, Madrid, Pearson, 2004].

confiárseles la toma de decisiones de importancia en salud, educación y justicia criminal. Cuando el único enfoque son las técnicas algorítmicas específicas, se está sugiriendo que solo importa el progreso técnico continuo, sin considerar el costo computacional de estos acercamientos ni sus impactos de alto alcance en un planeta bajo presión.

Por el contrario, en este libro sostengo que la IA no es *artificial* ni *inteligente*. Más bien existe de forma corpórea, como algo material, hecho de recursos naturales, combustible, mano de obra, infraestructuras, logística, historias y clasificaciones. Los sistemas de IA no son autónomos, racionales ni capaces de discernir algo sin un entrenamiento extenso y computacionalmente intensivo, con enormes conjuntos de datos o reglas y recompensas predefinidas. De hecho, la IA como la conocemos depende por completo de un conjunto mucho más vasto de estructuras políticas y sociales. Y, debido al capital que se necesita para construir IA a gran escala y a las maneras de ver que optimiza, los sistemas de IA son, al fin y al cabo, diseñados para servir a intereses dominantes ya existentes. En ese sentido, la IA es un certificado de poder.

En este libro, exploraremos cómo se hace la IA, en el sentido más amplio, y las fuerzas económicas, políticas, culturales e históricas que la moldean. Una vez que conectamos la IA con estas estructuras y sistemas sociales más extendidos, podemos escapar de la noción de que es un dominio puramente técnico. En un nivel más básico, la IA es una serie de prácticas técnicas y sociales; es instituciones e infraestructuras, política y cultura. El razonamiento computacional y el trabajo físico están profundamente enraizados: los sistemas de IA reflejan y producen al mismo tiempo relaciones sociales y entendimientos del mundo.

Vale la pena señalar que el término "inteligencia artificial" puede crear malestar en la comunidad informática. La frase se ha puesto y pasado de moda a lo largo de las décadas, y se usa

más en el *marketing* que en la investigación. En la bibliografía técnica, se emplea más a menudo la expresión "aprendizaje automático". Sin embargo, la nomenclatura de IA se adopta con frecuencia cuando se solicitan financiamientos, cuando los emprendedores aparecen con sus chequeras, o cuando los investigadores buscan llamar la atención de la prensa sobre un nuevo descubrimiento científico. Como consecuencia, el término es usado y rechazado a la vez, manteniendo su significado en constante cambio. Para los propósitos de este libro, uso el término IA para hablar sobre esa enorme formación industrial, que incluye política, mano de obra, cultura y capital. Cuando me refiero a aprendizaje automático, hablo de un rango de acercamientos técnicos (que también son, de hecho, de naturaleza social e infraestructural, aunque rara vez se los catalogue de esa manera).

Hay razones significativas que explican *por qué* el campo se ha concentrado tanto en el aspecto técnico –avances en los algoritmos, mejoras crecientes del producto y mayor conveniencia–. Las estructuras de poder que se ubican en la intersección entre la tecnología, el capital y el gobierno están bien cubiertas por este análisis limitado y abstracto. Pero, para entender que la IA es fundamentalmente política, debemos ir más allá de las redes neuronales y el reconocimiento de patrones estadísticos y preguntarnos, en cambio, *qué* se está optimizando, *para quién* y *quién* toma esas decisiones. Entonces, podemos rastrear sus implicaciones.

Ver la inteligencia artificial como un atlas

¿Cómo puede ayudarnos un atlas a entender cómo se hace la IA? Un atlas es un tipo de libro inusitado. Es una colección de partes dispares, con mapas que varían en resolución desde una

vista satelital del planeta a un detalle ampliado de un archipiélago. Cuando abres un atlas es posible que estés buscando información específica sobre un lugar en particular, o quizás estés errando, siguiendo tu curiosidad, encontrando senderos inesperados y nuevas perspectivas. Como señala la historiadora de la ciencia Lorraine Daston, todos los atlas científicos buscan entrenar el ojo, llevar la atención del observador a detalles particularmente reveladores y a características importantes.[19] El atlas te presenta un punto de vista específico del mundo, con el visto bueno de la ciencia (escalas y radios, latitudes y longitudes), pero también una sensación de forma y consistencia.

Y, sin embargo, un atlas no solo es un acto creativo –una intervención subjetiva, política y estética–, sino también una recopilación científica. El filósofo francés Georges Didi-Huberman piensa que el atlas es algo que habita tanto el paradigma estético de lo visual como el paradigma epistémico del conocimiento. Esta concomitancia socava la idea de que la ciencia y las artes están siempre completamente separadas.[20] Por el contrario, un atlas nos ofrece la posibilidad de releer el mundo, conectando las piezas de forma distinta y "reeditando y volviendo a armarlo sin pensar que lo estamos resumiendo o agotando".[21]

Quizá mi explicación favorita de cuán útil puede llegar a ser un acercamiento cartográfico la brinda Ursula Franklin, física y crítica de la tecnología: "Los mapas representan empeños intencionados: están destinados a ser útiles, a ayudar al viajero

[19] Lorraine Daston, "Cloud Physiognomy", en *Representations,* vol. 135, núm. 1, 2016, pp. 45-71, disponible en línea: <https://doi.org/10.1525/rep.2016.135.1.45>.

[20] Georges Didi-Huberman, *Atlas, or the Anxious Gay Science. How to Carry the World on One's Back?,* Chicago, University of Chicago Press, 2018, p. 5 [trad. esp.: *Atlas. ¿Cómo llevar el mundo a cuestas?,* Madrid, TF, 2010].

[21] *Ibid.,* p. 11.

cubriendo la brecha entre lo conocido y lo desconocido; son un testamento del conocimiento y la perspicacia colectivos".[22]

En el mejor de los casos, los mapas nos ofrecen un compendio de caminos abiertos –formas compartidas del conocimiento–, que pueden mezclarse y combinarse para crear nuevas interconexiones. Pero también existen los mapas de dominación, esos mapas nacionales en los que el territorio se ve tallado a lo largo de las fallas geológicas del poder: desde la intervención directa del trazado de fronteras sobre territorios en disputa hasta la revelación de los caminos coloniales de los imperios. Al invocar un atlas estoy sugiriendo que necesitamos nuevas maneras de entender los imperios de la IA. Necesitamos una teoría de la IA que tome en cuenta los Estados y las corporaciones que la manejan y dominan, la minería de extracción que deja una huella en el planeta, la captura masiva de datos y las prácticas laborales profundamente desiguales y cada vez más explotadoras que la sustentan. Estos son los movimientos tectónicos de poder en la IA. Un acercamiento topográfico ofrece distintas perspectivas y escalas, más allá de las promesas abstractas de la IA o de los últimos modelos de aprendizaje automático. La meta es entenderla en un contexto más amplio, atravesando los diferentes paisajes de la computación y viendo cómo se conectan.[23]

[22] Ursula M. Franklin y Michelle Swenarchuk, *The Ursula Franklin Reader. Pacifism as a Map,* Toronto, Between the Lines, 2006.

[23] Para un recuento de las prácticas de la colonización de datos, véanse Nick Couldry y Ulises A. Mejías, "Colonized by Data: The Costs of Connection with Nick Couldry and Ulises Mejías", charla, Berkman Klein Center for Internet and Society, Harvard University, 19 de septiembre de 2019, disponible en línea: <https://cyber.harvard.edu/events/colonized-data-costs-connection-nick-couldry-and-ulises-mejias>, y Achille Mbembe, *Critique of Black Reason,* Durham, Duke University Press, 2017 [trad. esp.: *Crítica de la razón negra. Ensayo sobre el racismo contemporáneo,* Barcelona, Futuro Anterior y Nuevos Emprendimientos Editoriales, 2016].

Los atlas también son relevantes aquí por otro motivo. El campo de la IA está explícitamente intentando capturar el planeta de una forma que sea legible por computadora. Esto no es una metáfora, sino la ambición directa de la industria, que está creando y normalizando sus propios mapas, a modo de visión cenital centralizada del movimiento, la comunicación y la mano de obra humana. Algunos científicos de la IA han declarado su deseo de capturar el mundo y de sustituir otras formas de conocimiento. La profesora de IA Fei-Fei Li describe que el objetivo de su proyecto ImageNet es "trazar un mapa de todo el mundo de los objetos".[24] En su manual, Russell y Norvig describen la IA como "relevante para cualquier tarea intelectual; es realmente un campo universal".[25] Woody Bledsoe, uno de los fundadores de la IA y de los primeros en hacer experimentos con el reconocimiento facial, lo dijo sin tapujos: "A largo plazo, la IA es la *única* ciencia".[26] Este no es un deseo de crear un atlas del mundo, sino de ser *el* atlas: la forma dominante de ver las cosas. Este impulso colonizador centraliza el poder en el campo de la IA: determina cómo se mide y define el mundo, mientras, al mismo tiempo, niega que se trate de una actividad inherentemente política.

En vez de fingir ser universal, este libro es un informe parcial. Al llevarlos conmigo a través de mis investigaciones, espero mostrarles cómo se fueron formando mis ideas. Nos encontraremos con los paisajes más reconocibles de la computación, así como con otros menos conocidos: las minas a cielo abierto, los largos pasi-

[24] Fei-Fei Li, cit. en Dave Gershgorn, "The Data That Transformed AI Research, and Possibly the World", en *Quartz,* 26 de julio de 2017, disponible en línea: <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/>.

[25] Stuart J. Russell y Peter Norvig, *op.cit.,* p. 1.

[26] Woodrow Wilson Bledsoe, cit. en Pamela McCorduck, *Machines Who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence,* Natick, A. K. Peters, 2004, p. 136.

llos de centros de datos que devoran cantidades inmensas de energía, los archivos de cráneos, las bases de datos de imágenes y los hangares con luces fluorescentes de los centros de distribución. Incluyo estos sitios no solo para ilustrar la construcción material de la IA y sus ideologías, sino también para "iluminar los aspectos inevitablemente subjetivos y políticos de la creación de mapas, y para proporcionar alternativas a los enfoques hegemónicos, autoritarios (y a menudo naturalizados y cosificados)", en palabras de Shannon Mattern, académica experta en medios.[27]

Los modelos para comprender y hacer rendir cuentas a los sistemas han reposado largamente en los ideales de la transparencia. Como he escrito junto con Mike Ananny, académico experto en medios, a veces se equipara el poder *ver* un sistema con el saber cómo funciona y cómo gobernarlo.[28] Pero esta tendencia tiene serias limitaciones. En el caso de la IA, no hay una única caja negra por abrir, ningún secreto por exponer, sino una multitud de sistemas de poder entrelazados. Por lo tanto, la transparencia completa es una meta inalcanzable. Mejor dicho, adquirimos un mejor entendimiento de la función de la IA en el mundo al enfrentarnos a su arquitectura material, sus entornos contextuales y políticas imperantes, y rastreando cómo se conectan entre sí.

Fundo mis opiniones en este libro en las disciplinas de estudios científicos y tecnológicos, en las leyes y la filosofía política, y en mi experiencia de trabajo tanto en la academia como en un laboratorio de investigación de IA industrial durante más de una década. A lo largo de esos años, la generosidad de muchos colegas y comunidades ha cambiado la manera en que veo el mundo:

[27] Shannon Mattern, *Code and Clay, Data and Dirt. Five Thousand Years of Urban Media,* Mineápolis, University of Minnesota Press, 2017, pp. 34 y 35.

[28] Mike Ananny y Kate Crawford, "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability", en *New Media and Society,* vol. 20, núm. 3, 2018, pp. 973-989.

crear mapas siempre es un ejercicio colectivo y esta no ha sido la excepción.[29] Agradezco a los académicos que crearon nuevas formas de entender los sistemas sociotécnicos, incluidos Geoffrey Bowker, Benjamin Bratton, Wendy Chun, Lorraine Daston, Peter Galison, Ian Hacking, Stuart Hall, Donald MacKenzie, Achille Mbembe, Alondra Nelson, Susan Leigh Star y Lucy Suchman, entre muchos otros. El presente ensayo se ha beneficiado de muchas conversaciones en persona, así como de la lectura de los trabajos recientes de autores que estudian las políticas de la tecnología: Mark Andrejevic, Ruha Benjamin, Meredith Broussard, Simone Browne, Julie Cohen, Sasha Costanza-Chock, Virginia Eubanks, Tarleton Gillespie, Mar Hicks, Tung-Hui Hu, Yuk Hui, Safiya Umoja Noble y Astra Taylor.

Al igual que cualquier otro libro, este surge de mi experiencia personal, lo que impone limitaciones. Como alguien que ha vivido y trabajado en Estados Unidos durante la última década, mi foco vira hacia la industria de la IA en los centros de poder occidentales. Sin embargo, mi objetivo no es el de crear un atlas global completo; esa sola idea invoca una captura y un control coloniales. Por el contrario, la visión de cualquier autor solo puede ser parcial, basada en observaciones e interpretaciones locales, en lo que la geógrafa ambiental Samantha Saville ha llamado una "geografía humilde" que da cuenta de la perspectiva específica propia antes que presumir de objetividad o dominio.[30]

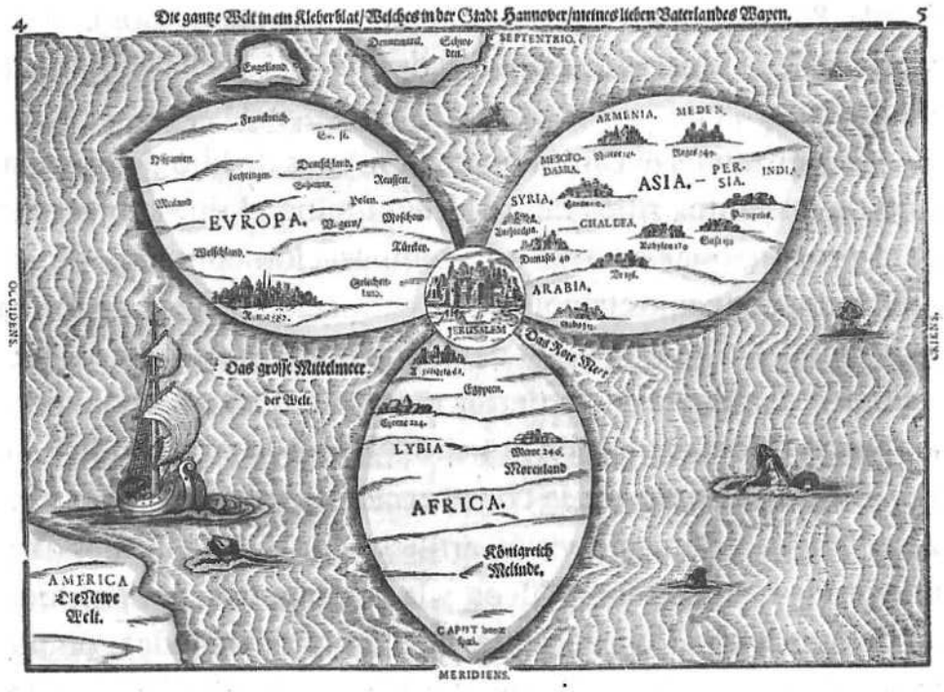
[29] Cualquier lista siempre será un recuento inadecuado de todas las personas y comunidades que han inspirado e informado este trabajo. Estoy particularmente agradecida de las siguientes comunidades de investigación: Fairness, Accountability, Transparency and Ethics (FATE) y Social Media Collective en Microsoft Research, el AI Now Institute de New York University, el grupo de trabajo Foundations of AI en la École Normale Supérieure, y el Richard von Weizsäcker Visiting Fellows en la Robert Bosch Academy en Berlín.

[30] Samantha Saville, "Towards Humble Geographies", en *Area,* 2019, pp. 1-9, disponible en línea: <https://doi.org/10.1111/area.12664>.

Así como hay muchas maneras de hacer un atlas, en el futuro la IA también podrá usarse de muchas formas distintas. El alcance expansivo de los sistemas de IA puede parecer inevitable, pero esta aseveración resulta discutible e incompleta. Los enfoques subyacentes del campo no surgen de forma autónoma, sino que, por el contrario, se han construido a partir de un conjunto de creencias y perspectivas específicas. Los principales diseñadores detrás del atlas contemporáneo de la IA pertenecen a un grupo pequeño y homogéneo de personas, radicado en un puñado de ciudades, que trabaja en una industria que actualmente es la más próspera del mundo. Como uno de esos mapamundis europeos medievales, que ilustraban tanto conceptos religiosos y clásicos como coordenadas geográficas, los mapas creados por la industria de la IA son intervenciones políticas, no imágenes neutrales del mundo. Este libro está hecho en oposición al espíritu de la lógica cartográfica colonial y abarca diferentes historias, locaciones y bases del conocimiento para comprender mejor la función de la IA en el mundo.

Topografías de la computación

¿Cómo se conceptualiza y construye la IA a estas alturas del siglo XXI? ¿Qué está en juego en el giro hacia la IA y qué tipos de políticas están contenidas en la manera en que se cartografía e interpreta el mundo? ¿Cuáles son las consecuencias, sociales y materiales, de incluir la IA y los sistemas algorítmicos vinculados a ella en los sistemas de toma de decisiones de instituciones sociales, como la educación y la atención médica, las finanzas, las operaciones gubernamentales, las interacciones en el lugar de trabajo y las contrataciones, los sistemas de comunicación y el sistema judicial? Este libro no es una historia acerca de



Mapamundi de Heinrich Bünting, conocido como "El mundo en una hoja de trébol", que simboliza la Santísima Trinidad, con la ciudad de Jerusalén como centro del mundo, rodeada por tres continentes centrales. Incluido en Itinerarium Sacrae Scripturae *(Magdeburgo, 1581).*

la programación y los algoritmos, ni acerca de las últimas consideraciones sobre la visión artificial o el procesamiento de lenguaje natural o el aprendizaje reforzado. Muchos otros libros tratan esos asuntos. Tampoco es un relato etnográfico sobre los efectos de la IA en la experiencia de trabajo, vivienda o salud de una comunidad específica; aunque ciertamente nos hacen falta más de esos.

En cambio, es una visión amplia de la IA como una *industria de extracción.* La creación de los sistemas de IA contemporáneos depende de la explotación de los recursos energéticos y minerales del planeta, de la mano de obra barata y los datos a gran

escala. Para observar este funcionamiento, realizaremos una serie de viajes a lugares que revelan de qué está hecha verdaderamente la IA.

Comenzamos el capítulo I en las minas de litio de Nevada, uno de los tantos sitios de extracción mineral que se necesita para proporcionar energía a la computación contemporánea. En la minería, vemos la política de extracción de la IA de forma más literal. Existe una vasta demanda por parte del sector tecnológico de minerales de tierras raras, petróleo y carbón, pero los costos de esa extracción nunca los asume la propia industria. Por el lado del *software,* la construcción de modelos para procesar lenguaje natural y visión artificial consume enormes cantidades de recursos energéticos, y la competencia por producir modelos más rápidos y eficientes ha impulsado métodos computacionalmente codiciosos que expanden la huella de carbono. Desde los últimos árboles en Malasia, cosechados para producir látex para los primeros cables submarinos transatlánticos, hasta el gigante lago artificial de residuos tóxicos en la Mongolia Interior, rastreamos los lugares de nacimiento, tanto ambientales como humanos, de las redes planetarias de computación y vemos cómo siguen terraformando el planeta.

El capítulo II muestra que la IA está hecha de mano de obra humana. Vemos a los trabajadores digitales a destajo, a los que se les paga una miseria, haciendo clic en microtareas para que los sistemas de datos parezcan más inteligentes de lo que realmente son.[31] Nuestra travesía nos llevará al interior de los centros de distribución de Amazon, donde los empleados deben seguir el

[31] Para leer más sobre los trabajadores de *crowdsourcing,* véanse Mary L. Gray y Siddharth Suri, *Ghost Work. How to Stop Silicon Valley from Building a New Global Underclass,* Boston, Houghton Mifflin Harcourt, 2019, y Sarah T. Roberts, *Behind the Screen. Content Moderation in the Shadows of Social Media,* New Haven, Yale University Press, 2019.

ritmo de las cadencias algorítmicas de un vasto imperio logístico, y, en Chicago, visitaremos a los trabajadores de la carne en las líneas de procesamiento, donde se diseccionan y preparan los cadáveres de animales para el consumo. Luego, escucharemos a los trabajadores protestar contra la forma en que los sistemas de IA aumentan la vigilancia y el control de sus jefes.

La mano de obra también es una historia sobre el tiempo. Coordinar las acciones de los seres humanos con los movimientos repetitivos de los robots y de las líneas de producción siempre ha implicado controlar cuerpos en el espacio y en el tiempo.[32] Desde la invención del cronómetro hasta el TrueTime de Google, el proceso de coordinar el tiempo estuvo en el centro de la gestión del lugar de trabajo. Las tecnologías de la IA requieren y crean a la vez las condiciones para unos mecanismos de gestión temporal cada vez más granulares y precisos. Coordinar el tiempo exige información cada vez más detallada acerca de qué está haciendo la gente, cómo y cuándo.

El capítulo III se centra en la función de los datos. Todo el material digital de acceso público —incluidos los datos personales o potencialmente dañinos— puede recolectarse para entrenar conjuntos de datos que luego se utilizan para producir modelos de IA. Hay conjuntos de datos gigantescos llenos de selfis, gestos con la mano, gente manejando, bebés llorando, conversaciones de grupos sobre noticias de la década de 1990; todo para mejorar algoritmos que realizan funciones como el reconocimiento facial, las predicciones de lenguaje y la detección de objetos. Cuando estas colecciones de datos ya no son vistas como efectos personales de la gente, sino meramente como *infraestructura,* se asume que el significado o contexto de una imagen o video es

[32] Jimena Canales, *A Tenth of a Second. A History,* Chicago, University of Chicago Press, 2010.

irrelevante. Más allá del serio problema de la privacidad y del capitalismo de vigilancia vigente, las prácticas actuales de trabajar con datos en la IA plantean profundas preocupaciones éticas, metodológicas y epistemológicas.[33]

¿Y cómo se usan todos estos datos? En el capítulo IV, observaremos las prácticas de clasificación en los sistemas de IA, la "maquinaria epistémica", según la socióloga Karin Knorr Cetina.[34] Veremos que los sistemas actuales utilizan etiquetas para predecir la identidad humana, por lo general, a partir de géneros binarios, categorías raciales catalogadas como esenciales y valoraciones problemáticas de personalidad y solvencia. Un signo remplazará un sistema, un *proxy* simbolizará la cosa real y se le pedirá a un modelo de juguete que sustituya la infinita complejidad de la subjetividad humana. Al observar cómo se hacen las clasificaciones, vemos de qué manera los diagramas técnicos refuerzan las jerarquías y magnifican la desigualdad. El aprendizaje automático nos presenta un régimen normativo que, cuando va en ascenso, adopta la forma de una racionalidad gobernante poderosa.

Desde aquí, viajamos a los pueblos serranos de Papúa Nueva Guinea para explorar la historia del reconocimiento emocional, que sostiene que la clave para revelar el estado emocional interior de una persona se encuentra en sus expresiones faciales. El capítulo V sopesa la aseveración del psicólogo Paul Ekman según la cual hay un pequeño conjunto de estados emocionales universales que se pueden leer directamente en el rostro. Las empresas de tecnología están ahora utilizando esta idea en sistemas de reconocimiento emocional, como parte de una industria que, se predice, estará valorada en más de 17.000 millones de dólares.[35] Pero hay una considerable controversia científica en torno a la detección de emociones, la cual en el mejor de los casos es incompleta y, en el peor, engañosa. A pesar de la inestable premisa, estas herramientas están siendo rápidamente utilizadas en los sistemas de contratación, educación y vigilancia.

[33] Shoshana Zuboff, *The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power,* Nueva York, PublicAffairs, 2019 [trad. esp.: *La era del capitalismo de la vigilancia. La lucha por un futuro humano frente a las nuevas fronteras del poder,* Buenos Aires, Paidós, 2021].

[34] Karin Knorr Cetina, *Epistemic Cultures. How the Sciences Make Knowledge,* Cambridge, Harvard University Press, 1999, p. 3.

En el capítulo VI, veremos los modos en que los sistemas de IA son usados como una herramienta del poder estatal. El pasado y el presente militar de la IA han moldeado las prácticas de vigilancia, extracción de datos y evaluación de riesgos que vemos hoy. Las profundas interconexiones entre el sector tecnológico y el ejército están siendo manipuladas para adaptarlas a una fuerte agenda nacionalista. Mientras tanto, las herramientas ilegales usadas por la comunidad de inteligencia se han dispersado y pasaron del mundo militar al sector tecnológico comercial, para ser usadas en aulas, estaciones de policía, lugares de trabajo y agencias de desempleo. La lógica militar que le dio forma a los sistemas de IA ahora es parte del funcionamiento de los gobiernos municipales y está sesgando aún más la relación entre los Estados y los sujetos.

El capítulo final evalúa la manera en que la IA actúa como una estructura de poder que combina infraestructura, capital y mano de obra. Desde la alerta que recibe el conductor de Uber, al rastreo del inmigrante indocumentado o a los inquilinos de viviendas sociales que lidian en sus hogares con sistemas de reconocimiento facial, los sistemas de IA se construyen siguiendo la lógica del capital, la vigilancia y la militarización. Esta combi-

[35] "Emotion Detection and Recognition (EDR) Market Size to Surpass 18%+ CAGR 2020 to 2027", en *MarketWatch,* 5 de octubre de 2020.

nación amplía más las asimetrías de poder ya existentes. Estas formas de ver dependen de los movimientos especulares de la abstracción y la extracción: abstraer las condiciones materiales de su creación mientras se extrae más información y recursos de aquellos que son menos capaces de resistirse.

Sin embargo, estas lógicas pueden ser cuestionadas del mismo modo en que los sistemas que perpetúan la opresión pueden ser rechazados. A medida que las condiciones en la Tierra cambian, los llamados a proteger datos, derechos laborales, justicia climática y equidad racial deberían ser escuchados en conjunto. Cuando entendemos la IA a través de estos movimientos interconectados en pos de la justicia, son posibles diferentes concepciones de políticas planetarias.

Extracción, poder y política

La IA es, por lo tanto, una idea, una infraestructura, una industria, una forma de ejercer poder y una manera de ver; también es la manifestación de un capital muy organizado respaldado por vastos sistemas de extracción y logística, con cadenas de suministro que abarcan todo el planeta. Todas estas cosas forman parte de la IA: una frase de dos palabras sobre la cual se puede cartografiar un conjunto complejo de expectativas, ideologías, deseos y miedos.

La IA puede parecer una fuerza espectral, un tipo de computación incorpórea, pero estos sistemas no son abstractos en absoluto. Son infraestructuras físicas que están transformando la Tierra, a la vez que alteran la forma en que vemos y entendemos el mundo.

Es importante que lidiemos con los múltiples aspectos de la IA: su maleabilidad, su desorden y su alcance espacial y temporal.

La promiscuidad de la IA como término, su apertura a ser reconfigurada, quiere decir que se la puede utilizar de diversas formas: se puede referir a cualquier cosa, desde dispositivos de consumo como el Amazon Echo hasta sistemas anónimos de procesamientos secundarios; desde artículos técnicos especializados hasta las compañías industriales más grandes del mundo. Pero esto también tiene sus ventajas. La amplitud del término "inteligencia artificial" nos permite considerar todos estos elementos y cómo están profundamente imbricados: desde la política de la inteligencia hasta la recolección masiva de datos; desde la concentración industrial del sector de la tecnología hasta el poder militar geopolítico; desde el entorno de desarraigo hasta las formas de discriminación en desarrollo.

Nuestra tarea es permanecer pendientes del terreno y observar los significados cambiantes y maleables del término "inteligencia artificial" —como un contenedor en el que se colocan varias cosas y luego se retiran—, porque eso también es parte de la historia.

En pocas palabras, la IA cumple ahora una función en la creación del conocimiento, la comunicación y el poder. Estas reconfiguraciones están ocurriendo a nivel de la epistemología, los principios de justicia, la organización social, la expresión política, la cultura, el entendimiento de los cuerpos humanos, las subjetividades y las identidades; lo que somos y lo que podemos ser. Pero es posible ir todavía más lejos. La IA no solo es una parte del proceso de replantear e intervenir el mundo, sino una forma fundamentalmente política de hacer el mundo, aunque rara vez se la reconozca como tal. Este proceso es no democrático, dominado por las grandes casas de la IA, que constan de media docena de compañías que dominan la computación planetaria a gran escala.

Muchas instituciones sociales están influidas por estas herramientas y métodos que dan forma a lo que valoran y a la manera

en que toman decisiones, mientras crean una compleja cadena de efectos posteriores. Durante largo tiempo, el poder tecnocrático se ha intensificado; sin embargo, el proceso ahora se ha acelerado. Esto se debe, en parte, a la concentración del capital industrial en una época de austeridad económica y tercerización, que ha incluido la eliminación de fondos para los sistemas de ayuda social y para las instituciones que alguna vez actuaron como reguladoras del poder del mercado. Por ello, debemos lidiar con la IA como una fuerza política, económica, cultural y científica. Como observan Alondra Nelson, Thuy Linh Tu y Alicia Headlam Hines: "Las controversias en torno a la tecnología están siempre relacionadas con luchas más grandes por la movilidad económica, la maniobrabilidad política y la construcción de comunidades".[36]

Nos encontramos en una coyuntura crítica que nos obliga a hacernos preguntas difíciles sobre la forma en que la IA se produce y adopta: ¿qué es la IA?, ¿qué tipos de políticas difunde?, ¿a qué intereses responde y quién corre el mayor riesgo de sufrir daños?, ¿adónde debería limitarse el uso de la IA? Estas preguntas no tendrán respuestas cómodas. No obstante, no es una situación sin solución o un punto sin retorno; el pensamiento apocalíptico nos puede paralizar hasta impedirnos tomar medidas o realizar intervenciones que se necesitan con urgencia.[37] Como ha escrito Ursula Franklin: "La viabilidad de la tecnología,

[36] Alondra Nelson, Thuy Linh N. Tu y Alicia Headlam Hines, "Introduction: Hidden Circuits", en Alondra Nelson, Thuy Linh N. Tu y Alicia Headlam Hines (eds.), *Technicolor. Race, Technology, and Everyday Life,* Nueva York, New York University Press, 2001, p. 5.

[37] Déborah Danowski y Eduardo Batalha Viveiros de Castro, *The Ends of the World,* trad. de Rodrigo Guimaraes Nunes, Malden, Polity, 2017 [trad. esp.: *¿Hay mundo por venir? Ensayo sobre los miedos y los fines,* trad. de Rodrigo Álvarez, Buenos Aires, Caja Negra, 2019].

como la democracia, depende a fin de cuentas de la práctica de la justicia y de la aplicación de límites al poder".[38]

Este libro sostiene que, para abordar los problemas fundacionales de la IA y la computación planetaria, primero debemos conectar los temas del poder y la justicia: de la epistemología a los derechos laborales, de la extracción de recursos a la protección de datos, de la desigualdad racial al cambio climático. Para hacer eso, necesitamos expandir nuestro entendimiento de lo que está ocurriendo en los imperios de la IA, para ver lo que está en juego y tomar mejores decisiones colectivas sobre lo que vendrá a continuación.

[38] Ursula M. Franklin, *The Real World of Technology,* ed. rev., Toronto, House of Anansi Press, 2004, p. 5.